Genetic and Environmental Factors Affecting The Human Gut Microbiom In Obesity

Beaumont, Michelle

Awarding institution:
King's College London

GENETIC AND ENVIRONMENTAL FACTORS AFFECTING THE HUMAN GUT MICROBIOME IN OBESITY

MICHELLE BEAUMONT

Doctor of Philosophy (PhD)

Department of Twin Research and Genetic Epidemiology
Division of Genetics and Molecular Medicine
King's College London

April 2016

SUPERVISORS:

Jordana Bell

Tim Spector

*A*ll our dreams can come true if we have the courage to pursue them.

— Walt Disney

ABSTRACT

The human microbiota is defined as the collective microbes living on and in the human body. Microbes can confer many benefits to the host including pathogen protection, complex carbohydrate breakdown and essential compound biosynthesis. Imbalances in the microbiome have been linked to disease, in particular obesity. The aim of this thesis is to explore the genetic and environmental effects on the human gut microbiome in twins, and characterise the microbiome species related to obesity. The primary dataset included gut microbiome 16s profiles of 982 twins from the TwinsUK cohort with extensive phenotype data.

I first explored the heritability of the gut microbiome and obesity using twin modelling. The results revealed several heritable families, including the most heritable family, Christensenellaceae, which was also associated with lean BMI and measures of adiposity, including visceral fat. In addition, I found further associations between *Christensenella* and immune-related phenotypes such as white blood cell count, as well as host immune genetic variants.

The next major analysis focused on identifying gut bacterial species that significantly differed between obese and lean individuals. I refer to these analyses as a Microbiome-Wide Association Study (MiWAS). The majority of the significant MiWAS results in obesity were obtained for members of the Ruminococcaceae and Lachnospiraceae. A candidate gene analysis was performed which aimed to identify obesity-associated human variants linked to these microbes. The strongest association was obtained between Lachnospiraceae and a variant within the human *RPTOR* gene, which controls the insulin-signalling path-

way in response to nutrient availability by either binding or dissociating from MTOR. This finding and further associations of Lachnospiraceae with insulin, diet and cholesterol measures implicate a role for Lachnospiraceae in insulin resistance as well as obesity.

The final project explored the association between the gut microbiome and metabolites in serum, plasma and faeces. First I identified associations between serum and plasma metabolites. The peak serum & plasma metabolite signals were obtained with the metabolites palmitate and cholesterol, which both play an important role in human metabolic health. The top associations were with the family Christensenellaceae indicating the importance of this microbial family not only in obesity, but also in extended metabolic phenotypes. Next, I obtained strong signals both between adiposity and faecal metabolites as well as with 16S microbial profiles. This indicates a useful role for faecal metabolomics as well as serum/plasma metabolomics in future phenotypic analyses.

In conclusion, my thesis explored the genetic and environmental basis of the human microbiome in twins, identifying host genetic influences on the gut microbiome. The project also aimed to characterise the human gut microbiome in obesity. The results confirm and extend previous findings of obesity-related microbiome profiles, and identify interesting host genes that may provide a starting point to understand the interaction of host genetics with the gut microbiome in human metabolism and obesity.

PUBLICATIONS RELATED TO THE THESIS

Some ideas and figures have appeared previously in the following publications:

[1] Goodrich, JK., Waters, JL., Poole, AC., Sutter, JL., Koren, O., Blekhman, R., **Beaumont, M.**, Van Treuren, W., Knight, R., Bell, JT., Spector, TD., Clark, AG., Ley, RE. Human Genetics Shape the Gut Microbiome. *Cell* 159, 789-799, (2014).

[2] Barrios, C., **Beaumont, M.**, Pallister, T., Villar, J., Goodrich, JK., Clark, AG., Pascual, J., Ley, RE., Spector, TD., Bell, JT., Menni, C. Gut-Microbiota-Metabolite Axis in Early Renal Function Decline. *PLoS ONE* 10(8), (2015).

To know, is to know that you know nothing.

That is the meaning of true knowledge.

— Socrates

ACKNOWLEDGEMENTS

Here I sit at 3am, staring at a blank page, inexplicably struck dumb at how difficult it actually is to find words that fully express my deep gratitude. In any case I'll try...

While this thesis is in my name only, a great many people have helped me to accomplish the completion of this work. First and foremost, I would like to express my warmest, heartfelt gratitude to my supervisors, Dr. Jordana Bell and Prof. Tim Spector. You have taught me more than I could have hoped to learn and I look forward to learning more still. I would also like to thank my various colleagues who I have worked (and socialised!) with over the last four years, both in the Department of Twin Research and the Ley Lab.

I would like to give special thanks to my "volunteer family" (and Andy!) at London Zoo as well as everyone on Mammals North. Despite snow, hail, heatwaves and floods you have all given me amazing experiences, friendships and at times a much needed respite from the crazy world of academia. Thank you!

Three people in particular deserve special mention. Amanda, my oldest friend, though you may not have been nearby (or even on the same side of the World!) during the last 4 years, you helped me in more ways than I can list here. I hope I can be just as helpful during your PhD. Rachel, thank you for listening to my rants, my insane ideas (and for going along with said ideas) and for being there whenever I've needed you. You're a star! And James, where do I possibly

begin?! There are literally no words to thank you for your unwavering support. Your obsession with cute things and random outbursts never fail to brighten my day and although I hate to admit it, the nickname "Slug" has grown on me.

None of this would have been possible if it weren't for one person: my Grandma. I will never be able to express just how grateful I am for your help and guidance throughout not just this PhD but also my life.

Finally, a big thank you to my Mum. Where would I be without you?

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

CD39 Cluster of Differentiation 39

CKD Chronic Kidney Disease

CT Computed Tomography

DEXA Dual Energy X-Ray Absorptiometry

DIO Diet-Induced Obesity

eGFR Estimated Glomerular Filtration Rate

ENTPD1 Ectonucleoside Triphosphate Diphosphohydrolase 1

FACS Fluorescence Activate Cell Sorting

FFQ Food Frequency Questionnaire

GWAS Genome-Wide Association Study

HMP Human Microbiome Project

IBD Inflammatory Bowel Disease

IBS Irritable Bowel Syndrome

ICC Intraclass Correlation

MAF Minor Allele Frequency

MDP Muramyl Dipeptide

OTU Operational Taxonomic Unit

PCA Principle Components Analysis

PCoA Principle Coordinates Analysis

QC Quality Control

QTL Quantitative Trait Loci

RR Relative Risk

SCFA Short Chain Fatty Acid

SEM Structural Equation Modelling

WBC White Blood Cell

Part I

BACKGROUND

1

INTRODUCTION

The complexity of the human microbiome and its effect on human health has been the focus of research in recent years. Composed of mainly bacteria and archaea, these microbes make up approximately 100 trillion cells and live in a complex symbiosis with the host, providing pathogen protection and carbohydrate processing in return for a safe environment and nutrients. It is believed a healthy human possesses up to 15,000 species across all body sites, including the oral cavity, gut (via faecal samples), skin and vagina, but other, perhaps more invasive, areas are being increasingly studied such as the colon, lung and conjuctiva. In this Chapter I will provide an overview of variation in the human microbiome and how dysbiosis, defined as a microbial imbalance either on or in the human body, in host microbial communities has been linked with disease.

1.1 SPATIAL VARIATION

The human body possesses a number of sites and niches that microbes can colonise, and the diversity and structure of these colonies can vary greatly from site to site. The first study to categorise the differences in microbial communities between body sites was Costello and colleagues in 2009 [Costello et al., 2009]. Microbial communities assayed from hundreds of samples, collected from various body sites including the gut, external auditory canal, nasal cavity, oral cavity and a selection of skin sites, clustered according to the body site sam-

pled during PCoA analysis which explained the majority of the variation in microbial community composition. Skin had the highest level of intrapersonal variation while the oral cavity had the lowest variability. Both the gut and the skin were approximately equal in terms of having high interpersonal variability, likely due to the persistent exposure to environmental variables. In terms of taxa, the forehead, nose and ear possessed similar abundant taxa, in particular Propionibacteriaceae, but these sites were very different to areas elsewhere on the skin such as the palm of the hand or the sole of the foot. A subsequent study also found the same trend in body site diversity, showing the oral cavity and gut to have the highest diversity [Huttenhower et al., 2012].

It has been established that an individual only shares approximately 17% of their microbes on their palms, and only 13% of bacteria are shared between unrelated hands [Fierer et al., 2008], showing that both interpersonal and intrapersonal variation in the human gut microbiome is vast.

1.2 TEMPORAL VARIATION

The structure of human microbiome is constantly shifting from day to day, reacting to the various environmental pressures imposed upon it. The earliest study to address temporal shifts in the microbiota was performed by Caporaso and colleagues in 2011 [Caporaso et al., 2011]. Principle coordinates analysis (PCoA) revealed that samples clustered according to body site in high concordance with the results observed in the Costello et al. [2009], and that this clustering was highly stable over time while still showing variation within body sites. Furthermore, microbes in each person could either be classified as transient, being acquired and lost over short periods of time,

or persistent, remaining in the community for long periods of time but not long enough to be considered a core microbiota member. A later study [Flores et al., 2014]) revealed that this temporal variation differs dependent upon the body site. Forehead and palm sites possessed higher temporal variation in alpha diversity, and this variation appeared to be driven by the acquisition of transient taxa. Given the exposure of these two sites to constant environmental pressures, it is perhaps unsurprising to observe a higher level of variation over time here. Temporal variation within internal sites such as the gut and tongue however, appeared to be driven by the shifts in relative abundance of species present and it was impossible to use the variability of one site to predict the variability of another.

The microbiome has been shown to fluctuate with the seasons [Davenport et al., 2014]. The Hutterites are an isolated population of communal emigrants from Europe, making them an ideal population for study due to less heterogeneity in environmental factors and lower genetic diversity. They consume three meals a day in a communal setting, using fresh produce they grow themselves and recipes that have changed little in the 150 years since settling in America. It was in this population of people that Davenport et al (2014) found distinct shifts in microbiota composition from Summer to Winter and this is likely due to the availability of produce at certain times of the year. As reported in the paper, in the Summer, food containing high levels of complex carbohydrates were consumed, potentially explaining the increase in Bacteroidetes. Members of the Bacteroidetes possess a large number of carbohydrate processing genes [Flint et al., 2012] increasing the probability that this family of microbes is important for the degradation of complex carbohydrates.

1.3 CORE MICROBIOME AND HOST GENETICS

The subject of a core microbiome has been the centre of some debate since conception of the idea. The core microbiome theory postulates that some microbes are found in everyone, regardless of relatedness or environmental pressure because of the essential role they perform for the host [Turnbaugh et al., 2007]. In murine mouse models, a core gut microbiome of approximately 64 taxonomic species was defined [Benson et al., 2010] and these accounted for over 90% of the overall sequence reads. An additional human study found 35% of reads from the gut could be mapped to other samples, suggesting the presence of a core microbiome [Qin et al., 2010]. Yet as the skin shows less than 13% similarity between unrelated individuals, this raises the question of a difference in core communities across sites. The Human Microbiome Project (HMP) sought to define a core microbiome across individuals and sites [Huse et al., 2012]. In order to define a core microbiome, researchers set a threshold requiring operational taxonomic units (OTUs), that is closely related sequences that cluster together representing a taxonomic unit, to be present in more than 95% of samples for a given site to be considered a core OTU. At this threshold, the oral cavity displayed the largest core group, possessing between 7 and 22 core OTUs dependent upon the site sampled, and there were 4 OTUs shared across all oral sites. The vagina displayed the lowest number of core OTUs (n=3). Only one OTU was present across all skin sites (*Propionibacterium*) and no OTU was present across all sites and individuals. In concordance with the Benson et al study [Benson et al., 2010], the most abundant OTUs also tended to be the most prevalent.

Caporaso et al however, discovered little evidence for a core microbiome over time, at multiple thresholds for defining core microbiome

membership [Caporaso et al., 2011]. The cut-off that determines what should be considered a core OTU is subjective and indeed, whether we should be defining a core microbiome by OTU membership is subject to debate. While consensus about a taxonomic core microbiome is lacking, it does appear that there is a 'functional' core. Using approximately 150 obese and lean twins and their mothers, Turnbaugh et al. [2009a] discovered that while microbial profiles changed across people and time and displayed high levels of β-diversity, their gene profile, and thus predicted functional profile, remained highly stable. This was later confirmed in the larger HMP dataset using metagenomics [Huttenhower et al., 2012]. It is perhaps best to consider the microbiome in terms of its functional relevance. This is demonstrated in obesity and IBD where use of metagenomic sequencing and KEGG annotation revealed an enrichment of genes involved in membrane transport [Greenblum et al., 2012]. This may suggest an alteration in how obese microbiomes interact with the host. Another earlier study in mice revealed the obese microbiome has an enhanced capacity to harvest energy from consumed food [Turnbaugh et al., 2006] while in adolescent humans, an obese microbiome showed an increased ability to synthesise butyrate, an essential colonic energy source [Ferrer et al., 2013].

While the presence of a core microbiome may continue to be debated, the idea that there may be such an essential group of microbes naturally leads one to question whether or not this 'core' or indeed the presence of any other microbes, may be driven by host genetics. Twin studies can provide a simple model for dissecting the host genetic and environmental influences underlying a phenotype, such as the presence or relative abundance of a microbe. It is assumed that if a trait were caused by host genetics, then monozygotic twins would have more similar trait values, compared to DZ twins because

MZ twins share nearly identical genomes while DZ twins on average share only 50% of their genetic variants. However the assumption here is that environmental effects for the two types of twins, as well as between twins within a pair, are the same, although this is not necessarily correct.

A number of studies have demonstrated that microbiomes are more similar between monozygotic twins versus dizygotic twins [Zoetendal et al., 2001, Stewart et al., 2005, Turnbaugh et al., 2008]. These findings suggest that there may be a host-genetic component driving the microbiota. However, our recent large-scale study using almost 1000 twins performed an extensive heritability analysis to determine the contribution of host-genetics to every single OTU found within the sample [Goodrich et al., 2014]. I will discuss this in more detail in Chapter 3 where I outline my own contributions to this paper.

Genetic studies in the human host often require large numbers of study subjects to find effects associated with common genetic variants because the majority of common genetic effects tend to have moderate to modest effect sizes on human traits. The latest obesity genome-wide association study (GWAS) was a meta-analysis with a total study size of 339,224 individuals [Locke et al., 2015]. This larger study was able to detect new loci with a lower minor allele frequency (MAF) than previously found. However the loci discovered to date still only explain 2.7% of the variance seen in obesity, despite the obesity's high heritability. This demonstrates the difficulty of finding genetic effects in small populations unless the effect size is quite large. With microbiome studies being mostly limited in sample size, many using only a couple of hundred individuals, GWAS approaches tend to be extremely underpowered to detect host genetic variants that associate with microbes. Candidate gene approaches have had some success with microbiota studies however, particularly in mouse mod-

els. A number of genes have been found to be associated with changes in the microbiota including *Apoa1* [Zhang et al., 2010], *Nod2* [Frank et al., 2011, Natividad et al., 2012] and *Nlrp6* [Anand et al., 2012]. As over 90% of the mouse genome is conserved in humans, mice are good model organisms for studying the relationship between host genetics and microbiota.

1.4 INFLUENCE OF THE ENVIRONMENT ON THE MICROBIOME

The environment is the largest factor affecting the human microbiome. The number of pressures it can apply is vast and we see the results of this in the high level of interpersonal variation in microbiotas.

1.4.1 *Diet*

Diet is perhaps the most significant environmental pressure on the human gut microbiome and there have been a number of studies outlining the implications that diet may have upon gut bacteria. Mouse models have shown that consumption of a high fat diet leads to changes in the colonic bacteria [Hildebrandt et al., 2009, Turnbaugh et al., 2009b] and another, more recent study has revealed that this alteration may influence mood and depression [Bruce-Keller et al., 2015]. Dietary fat is not the only aspect of diet that can alter the gut microbiota. Reductions in the amount of carbohydrate consumed has a resultant effect on the levels of butyrate produced in the gut [Duncan et al., 2007], perhaps indicating a less acidic environment in the gut that may not be suitable for short chain fatty acid (SCFA) fermentation. Vegetarianism also has a profound effect upon the gut microbes [David et al., 2014, Zimmer et al., 2012], with lower levels in *Bacteroides* observed in vegan and vegetarian samples. Dietary fi-

bre has been studied extensively in microbiome research due to the benefits to both colonic health and microbial communities. Consumption of dietary fibre results in an increase of butyrate and other SCFA, and also leads to the increase of bacteria considered to be beneficial for health, such as *Bifidobacterium* and *Lactobacilli* [Shen et al., 2012, Arora et al., 2012] and the effects of this have been linked to allergy [Trompette et al., 2014], IBD [Thibault et al., 2010] and even autism [Macfabe, 2012]. Dietary interventions are the best method to observe changes in the human microbiota, and further studies are required to fully tease apart the underlying mechanisms that drive dietary effects upon the gut microbiome.

1.4.2 *Co-habitation*

Diet, while perhaps the most influential factor, is not the only environmental pressure that can influence the human microbiota. Living with other people in the same household has an effect upon your microbial communities, as does in fact, living with a dog [Song et al., 2013]. In this study it was found that couples living together shared more microbes than individuals living in different households. In addition, if there was a pet dog living in that household then the microbiome of the owners would be more similar to their own dog than to dogs in other households.

1.4.3 *Smoking*

There is some evidence to suggest that smoking affects the intestinal microbiota. Smoking cessation results in an increase in overall microbial diversity as well as an increase in the relative abundance of Firmicutes and Actinobacteria [Biedermann et al., 2013]. Despite

the interventional nature of the study, the sample size (10 subjects) was quite small and it remains the only published study to assess the impact of smoking on the gut microbiome. There is however further evidence to suggest that alterations to both the respiratory tract [Morris et al., 2013] and oral microbiotas [Mason et al., 2015] are linked with cigarette smoking.

1.4.4 *Method of Delivery*

Method of delivery, that is whether or not a baby is born vaginally or by caesarian section, is thought to affect the human microbial communities. During a natural, vaginal birth, the foetus is forced through the birthing canal, receiving essential microbes from both the canal and from any faeces or other matter surrounding the vagina. During caesarian section however this does not happen and the baby does not receive this initial dose of microbes. 16s profiling of mothers and babies delivered naturally or by caesarian section revealed large differences in the microbiomes of babies born by the different methods, particularly with the levels of *Lactobacillus* [Dominguez-Bello et al., 2010]. Furthermore, babies delivered via caesarian were no more similar to their mothers than any other person. These pertubations can take up to 6 months to stabilise into a 'normal' gut profile [Gronlund et al., 1999] although another study has suggested differences may be observed as late as seven years [Salminen et al., 2004]. If these differences do persist into late childhood, it is as yet unclear whether this dysbiosis would have any effects on the host health. Studies have shown that children who are born by caesarian section are at more risk of developing health issues such as allergy [Negele et al., 2004, Eggesbo et al., 2003] and obesity [Huh et al., 2012], but there are a

number of factors that may contribute to these associations that may not necessarily be linked with the microbiome.

1.5 DISEASE AND DYSBIOSIS

Dysbiosis in the human gut microbiome has so far been linked with a wide range of conditions and illnesses including obesity, IBD & IBS, cancer, autism and allergy, to name a few. Given how the host relies on the microbiome for essential functions, it is not altogether unreasonable to assume that imbalances in the microbiome can result in disease.

1.5.1 *Cancer*

A very large observational study (n=3,112,624) in 2008 reported an increased relative risk (RR) of a number of cancers with antibiotic use [Kilkkinen et al., 2008], including the most common cancers; breast, lung, prostate and colon. The observational nature of the study does not allow for adjustment for confounding factors, such as smoking which is a known risk factor for cancer, as this data was not collected. However later studies have confirmed links between antibiotic use and colorectal cancer [Wang et al., 2014], breast cancer [Sergentanis et al., 2010, Tamim et al., 2008] and prostate cancer [Tamim et al., 2010].

A number of studies have linked individual microbes to cancer. *E.coli* has been found to be significantly higher in individuals suffering from colorectal cancer than controls. When mice received *E.coli* orally, those receiving *E.coli* developed more colonic polyps than controls [Bonnet et al., 2014]. It has been suggested that butyrate and other SCFA produced by the colonic microbiota help to reduce col-

orectal cancer risk [O'Keefe et al., 2009], potentially by reducing inflammation [Arpaia et al., 2013]. This could perhaps explain the link between low-fat-high-fibre diets and breast cancer [Gorbach, 1984, Thomson, 2012], given that high fibre diets help to produce SCFA, while oestrogen, a known causative risk factor for breast cancer, can be metabolised from cholesterol. Breast cancer patients have also been shown to have a lower bacterial diversity in the gut, independent of oestrogen levels [Goedert et al., 2015].

1.5.2 *Inflammatory Bowel Disease*

Inflammatory Bowel Disease (IBD) is not just one disease, but a group of inflammatory disorders of the gut, principally Crohn's disease and ulcerative colitis. In Crohn's patients it was found that in the gut both Bacteroidetes and Proteobacteria were more abundant while *Clostridia* (of the Firmicutes phylum) were reduced [Gophna et al., 2006, Frank et al., 2007]. While the findings are not any indication of causation, the reduced abundance of *Clostridia* may have potential implications on the production of SCFA including butyrate, thus leading to inflammation.

Perhaps most intriguingly with Crohn's disease in particular is the effect mutations in the gene *NOD2* may have upon the gut microbiome. *NOD2* is a known Crohn's disease susceptibility gene [Ogura et al., 2001], and acts a sensor for the bacterial cell wall peptidoglycan, muramyl dipeptide (MDP) [Girardin et al., 2003]. *NOD2* is primarily expressed in the Paneth cells in the intestinal ileum [Stappenbeck et al., 2002] and thus likely to be important for host defence. Presence of *NOD2* mutations have been associated with changes in the human gut microbiota, particularly a decrease in *Faecalibacterium* and *Escherichia* [Frank et al., 2011]. *Nod2*-deficient mice show a re-

duced ability to prevent colonisation of pathogenic bacterial species [Petnicki-Ocwieja et al., 2009] and also possess higher loads of both Firmicutes and Bacteroidetes in the ileum [Rehman et al., 2011].

Probiotic treatment for Crohn's disease has had variable results, with some studies showing no difference between the case and control groups and others showing an effective difference [Saez-Lara et al., 2015]. This could perhaps be due to host genetic influence of *NOD2* or it may be down to the different strains used in each study. In ulcerative colitis however, probiotics seem to have a good effect overall in reducing the symptoms experienced by the host [Saez-Lara et al., 2015], providing a potential treatment option for sufferers.

1.5.3 *Diabetes*

Diabetes can be defined as two conditions: type 1 diabetes (T1D) and type 2 diabetes (T2D). T1D is characterised as the lack of insulin caused by the autoimmune destruction of pancreatic beta cells and patients require regular insulin injections to keep their glucose levels within the normal range. The cause of T1D remains unknown and it contributes to between 5-10% of all diabetes cases [Daneman, 2006]. T2D is characterised as insulin resistance leading to hyperglycaemia. High glucose and fatty acid levels lead to the phosphorylation of the IRS-1 receptor, and thus its impairment. This receptor no longer binds to insulin and glucose accumulates in the blood. A mixture of lifestyle factors and host genetics play a role in T2D but it can be managed through dietary adjustment and reduction of any cardiovascular risk factors the patient may have, for example, obesity.

Only 50% of monozygotic twins will show concordance in presenting T1D and in dizygotic twins the concordance rate dips to 10% [Achenbach et al., 2005]. This suggests that extraneous environmen-

tal factors must be responsible in part for triggering the disease. Research that followed infants predisposed to T1D found a marked change in microbial alpha diversity of infant guts that went on to develop T1D, more specifically, the alpha diversity changed in the window between seroconversion and T1D diagnosis [Kostic et al., 2015]. Diabetes-induced rats that were fed dahi, an Indian fermented dairy product containing *Lactobacillus acidophilus* and *Lactobacillus casei*, showed delayed onset of diabetic symptoms [Yadav et al., 2007]. Furthermore, germ-free NOD mice will develop diabetes however, NOD mice raised in specific pathogen free conditions will not present with diabetes, suggesting that key components of the microbiome may act as a trigger for the progression of T1D [Wen et al., 2008].

1.5.4 *Obesity*

Obesity is one of the most well-studied diseases in terms of microbiome dysbiosis. Mouse models were used to discover the initial link between obesity and the microbiota. Taking germ-free mice and inoculating them with the microbiota of conventionally raised mice led to a 60% increase in body fat of the germ-free mice [Backhed et al., 2004]. Furthermore, genetically obese mice revealed an increase in Firmicutes members compared with lean littermates [Ley et al., 2005]. This increase in Firmicutes was observed in a number of studies including in humans [Ley et al., 2006] however not all studies found this trend [Duncan et al., 2008], and one study even reported the opposite [Schwiertz et al., 2010]. A meta-analysis of obesity studies concluded that there were no consistent microbial signatures of obesity across studies [Walters et al., 2014]. Consumption of a fat or carbohydrate-restricted diet that resulted in weight loss also altered the ratio of Firmicutes to Bacteroidetes [Ley et al., 2006]. This

so-called Bacteroidetes:Firmicutes ratio imbalance between obese and lean individuals was found to favour energy harvest in obese subjects [Turnbaugh et al., 2006] suggesting that the dysbiosis observed in obese individuals helped them to extract more energy from food. Diet-induced obesity (DIO) was also transmissable by inoculating germ-free mice with the gut microbiomes of DIO mice [Turnbaugh et al., 2008] suggesting a causative role for the microbiome in the progression of obesity.

Research has begun to move towards pinpointing particular microbes that may influence obesity. When *Bacteroides thetaiotaomicron* cocolonises with *Methanobacteria brevii*, adiposity in the host increases, potentially due to the ability of this pairing to ferment dietary fructans [Samuel and Gordon, 2006]. In our collaborative study, it was observed that *M. brevii* co-occurs with *Christensenella minuta* and when mice were inoculated with twin microbiomes containing *C. minuta*, those receiving *C. minuta* gained less weight than those who did not receive it [Goodrich et al., 2014]. More detail on these analyses can be found in Chapter 3 and Chapter 4. Strains of *Lactobacillus* and *Bifidobacterium* may have a beneficial effect on adiposity reduction and weight gain [Mekkes et al., 2014, Wang et al., 2015] providing a promising probiotic weight-loss treatment option to explore.

Obese individuals often have reduced gut microbial alpha diversity [Goodrich et al., 2014, Turnbaugh and Gordon, 2009] showing that not only are the types of bacteria present in the gut different during obesity, but that there are fewer species in evidence. This may be an indication of the dietary and behavioural differences between lean and obese individuals, given that isolated hunter-gatherer communities such as the Amerindians and Hadza tribes have the most diverse microbial communities of any group of humans studied [Clemente

et al., 2015, Schnorr et al., 2014] while also possessing active lifestyles and a diet that is diverse and varies with seasonal availability.

1.6 AIMS

The aim of this research is to explore genetic and environmental influences on the gut microbiome in twins, to characterize the human gut microbiome in obesity, and to relate these findings to metabolomic profiles. To this end, I first applied twin modelling approaches to assess gut microbiome heritability and explore microbial associations with specific enviromental factors in Chapter 3. I then pursued large-scale analyses of the gut microbial profiles in twins with respect to obesity, adiposity, and related phenotypes in Chapter 4 and Chapter 5. Visceral fat is traditionally difficult to measure and as such, human microbiome analyses with visceral fat have yet to be performed. I addressed visceral fat and the microbiome in this thesis. In addition I explored the hypothesis that host genetics may influence the observed microbial associations in obesity. Lastly, I tested for association between gut microbial profiles and human blood and faecal metabolomics datasets, and explored the findings in the context of kidney function and obesity in Chapter 6.

Part II

RESEARCH

2

DATA AND METHODS

2.1 DATA

There are a number of subsets used within this body of research for each of the different analyses. This section will describe each dataset and the methods used to collect this data. Table 1 provides an overview of each dataset used in each thesis section and subsection, including sample sizes and zygosity and sex breakdowns.

	Section	Sample Size	Zygosity			Repeats	Sex	
			MZ	DZ	Singletons		Male	Female
	Full 16S Dataset	1081	342	490	145	104	20	1061
Subsets								
Chapter 3	Section 3.2.1 Heritability of the Gut Microbiota	1081	342	490	145	104	20	1061
	Section 3.2.2.1 Diet	838	278	422	138	0	0	838
	Section 3.2.2.2 Method of Delivery	536	124	328	84	0	0	536
	Section 3.2.2.3 Smoking	851	268	374	209	0	17	834
Chapter 4	Section 4.2.1 Christensenella vs Adiposity	960	344	474	142	0	19	941
	Section 4.2.2 Christensenella vs Inflammation	913	349	441	123	0	3	910
	Section 4.2.3.1 Adiposity Candidate Gene Analysis	886	320	440	126	0	13	873
	Section 4.2.3.2 Immune Candidate Gene Analysis	886	320	440	126	0	13	873
Chapter 5	Section 5.2.1 Heritability of Adiposity	726	310	416	0	0	14	712
	Section 5.2.2 Diversity vs Adiposity	960	344	474	142	0	19	941
	Section 5.2.3 MiWAS Obesity	960	344	474	142	0	19	941
	Section 5.2.4 Candidate Gene Analysis	886	320	440	126	0	13	873
Chapter 6	Section 6.2.1 Metabolite-wide, microbiome wide AS	831	272	428	131	0	10	821
	Section 6.2.4 Kidney Function with Microbiome	854	288	414	152	0	15	839
	Section 6.2.5.1 Faecal Metabolomics with Obesity	49	42	0	7	0	0	49
	Section 6.2.5.2 Faecal Metabolomics with 16S	27	22	0	5	0	0	1

Table 1: Description of all datasets used within this thesis, arranged by chapter.

2.1.1 *Diet Data*

Diet data in the TwinsUK cohort was previously collected, described and validated [Teucher et al., 2007]. Participants were asked to complete a 131-item food frequency questionnaire (FFQ), from which nu-

trients were derived and food items placed into groups based on similarity in nutrient content. Each food group had a frequency of intake calculated (Σ servings/week) that was then adjusted for total energy intake, and the resultant residuals where subsequently analysed using Principle Coordinates Analysis (PCoA). The first 5 PCs were kept for further analysis due to their representation of actual dietary patterns and explain 22% of the total variance (Table 2). The FFQs were conducted at two time points, and an average of the PCs was taken for further analysis.

Diet PC	Dietary Pattern	Description
Score 1	Fruit and Vegetable	Frequent intakes of fruit, allium and cruciferous vegetables; low intakes of fried potatoes.
Score 2	High Alcohol	Frequent intakes of beer, wine and allium vegetables; low intakes of high fiber breakfast cereals and fruit.
Score 3	Traditional English	Frequent intakes of fried fish and potatoes, meats, savoury pies and cruciferous vegetables.
Score 4	Dieting	Frequent intakes of low-fat dairy products; low-sugar soda, low intake of butter and sweet baked products.
Score 5	Low Meat	Frequent intakes of baked beans, pizza and soy foods; low intakes of meat, other fish, seafood and poulty.

Table 2: Dietary patterns in the TwinsUK cohort.

2.1.2 *Method of Birth Data*

Method of birth is collected from every subject via the first questionnaire they complete upon registering with the TwinsUK cohort. Rate of caesarian section in this cohort is approximately 4% which is much lower than the current national average of just over 50% (http://www.nhs.uk/conditions/pregnancy-and-baby/pages/giving-birth-to-twins.aspx), however much more representative of the time these twins were born around the 1950's

(http://www.parliament.uk/documents/post/pn184.pdf). Individuals that were unsure of their delivery method were removed from the dataset.

2.1.3 *Smoking Data*

Smoking data was collected via a number of questionnaires between 1992 and 2010. Subjects may indicate whether they are a current smoker, ex-smoker or have never smoked. A final status column was then created to indicate an overall status for each subject. If a subject consistently aligned to smoker, ex-smoker or never smoker, they were given a "clean" code of A, B or C. Subjects that answered inconsistently, but where the majority of responses aligned to one particular status, were given a provisional code of A4, B4 or C4. For my analyses I recoded these individuals to 0, 1 and 2 (Table 3).

Smoker	Designation	Definition	Inclusion of indeterminate answers	Coding	Analysis Coding
Never	Never Smoker	Responded only ever as a never/non-smoker or NA in all questions asked.	No	A	0
			Yes	A4	0
Ever	Forever Smoker	Responded only ever as a current smoker or NA in all questions asked.	No	B	1
			Yes	B4	1
	Ex-Smoker	Responded as a smoker, followed by responses as an ex/non smoker. NA included.	No	C	2
			Yes	C4	2
	Variable Status Smoker	Responded as both smoker and ex-smoker in no clear pattern. NA included.	No	D	2
			Yes	D4	2

Table 3: Smoking coding in the TwinsUK cohort and analyses with 16S data.

2.1.4 *Adiposity Data*

Adiposity data is collected during a subject's annual clinical visit. Visceral fat, % Trunk Fat, Android and Gynoid fat, lean and total mass as well as subcutaneous fat were estimated using a Dual Energy X-ray Absorptiometry (DEXA) machine, provided by Hologic.

Due to the different absorption properties of the tissues within the body, body composition can be estimated by measuring the absorption of x-rays at two different energies. This provides a non-invasive, accurate and direct measurement of adiposity [Hill et al., 2007]. Subjects were asked to lie flat, still and straight while a whole-body scan took place. Visceral fat mass was calculated from one cross-section of the whole body at L4-L5, the typical location of a CT slice. Height and weight measurements are collected during the annual visit also, and BMI was calculated as standard [Eknoyan, 2008]. Android:Gynoid ratio was calculated as android fat divided by gynoid fat.

2.1.5 *Blood Cell Counts and Insulin*

Blood cell counts, including white blood cell count (WBC), neutrophils, monocytes, eosinophils, and lymphocytes were measured in blood samples taken during a subject's annual clinic visit by the departmental laboratory using fluorescence activated cell sorting (FACS). Total white blood cell count was measured as thousands of cells per ml and subsequent cell sub-type counts were calculated by multiplying the proportion of the WBC count comprised by each cell type by the total WBC measure. This data has previously been described in Nalls et al (2011) *PLoS Genetics* [Nalls et al., 2011].

Fasting insulin (mIU/L) was measured by immunoassay (Abbott Laboratories Ltd.,Maidenhead, U.K.) as described in Jamshidi et al. [2006].

2.1.6 *Kidney Function Data*

Estimated glomerular filtration rate (eGFR, mL/min/1.73 m^2) measures the clearance of creatine from the blood and was estimated from

standard blood creatinine using the CKD-EPI equation [Levey et al., 2009]. Chronic Kidney Disease (CKD) outcome was defined as an eGFR <60 ml/min/1.73 m² according to the current Kidney Disease Outcome Quality Initiative (K/DOQI) guidelines. Both phenotypes have been previously described in Goek et al. [2012].

2.1.7 *Faecal 16S Profiles and Diversity*

Subjects were asked to collect a faecal sample at home and bring it to their annual clinic visit. Samples were collected in a 15mL conical tube and were refrigerated for approximately 1-2 days prior to the clinical visit. Recent literature has shown that storage of faecal samples at 4 °C for 3 days prior to freezing at −80 °C shows no significant alterations to the microbial profiles [Choo et al., 2015]. Once received, samples were frozen and stored at −80 °C before being shipped to Cornell University on dry ice. The following methods in this subsection were performed by Julia Goodrich in the Ley Lab at Cornell University.

DNA was extracted from the 100mg of faecal sample using the MoBio Power Soil® DNA isolation kit (MoBio Laboratories Ltd, Carlsbad, CA). PCR was performed using the primers 515F and 806R for the V4 hypervariable region of the 16S rRNA gene as previously described [Caporaso et al., 2011]. PCR was carried out in duplicate using 2.5 2.5 U Easy-A high-fidelity enzyme, 1 x buffer (Stratagene, La Jolla, CA), 10-100 ng DNA template, and 0.05 mM of each primer. The PCR cycles were as follows: initial denaturation at 94 °C for 3 minutes followed by 25 cycles of denaturation at 94 °C for 45 seconds, annealing at 50 °C for 60 seconds, extension at 72 °C for 90 seconds, and a final extension at 72 °C for 10 minutes. The PCR duplicates were combined and purified using the Mag-Bind® magnetic bead purifica-

tion system (EZPure, Omega Bio-Tek, Norcross, GA) then quantified using the QuantiT PicoGreen dsDNA assay kit (Invitrogen, Carlsbad, CA). Aliquots were combined to yield a final concentration of approximately 15 ng/µl.

Paired-end sequencing was performed on the Illumina MiSeq platform at Cornell Biotechnology Resource Center Genomics Facility. fast-q join (ea-utils package) was used to merge matched mate pair reads and reads exceeding 275bp in length were discarded. Sequences containing unreadable barcodes, ambiguous bases or reads with a Phred quality score $\leqslant 25$ were also discarded. This resulted in 81,634,331 quality-filtered seuqences. QIIME 1.7.0 (Quantitative Insights Into Microbial Ecology) was used to analyse the remaining sequences. Closed-reference OTU picking was used to pick OTUs at 97% sequence identity, against the May 2013 Greengenes database and the taxonomic assignment given to the reference sequence was given to each OTU which resulted in 9839 OTUs. Diversity analyses used the Greengenes phylogenetic tree and β diversity was calculated using an OTU table that had been rarified to 10,000 sequences per sample. Taxonomic summaries of the OTUs were generated at six levels from genus to phylum from the non-rarified OTU table. OTUs that were not present in over 50% of the samples were further excluded. The final OTU table was composed of 768 OTUs and 141 taxonomic summaries. This table was then adjusted for the following covariates using a linear model: identity of technician (of two), sequencing run (16 instrument runs), shipment batch (8 shipments), number of sequences per sample and age.

2.1.8 *Genotypes*

Genotypes were derived from an unselected population of 5654 subjects prior to this research. The genotyping and imputation steps for the TwinsUK cohort have been described in detail previously [Andrew et al., 2006, Zhai et al., 2011, Mangino et al., 2009]. Briefly, genoyping of the TwinsUK cohort was performed using a combination of Illumina arrays (HumanHap300, HumanHap610Q, 1M-Duo and 1.2MDuo). A number of quality control steps were previously applied to these genotype data by other researchers in the department, including genotype call posterior probability threshold greater than 0.95, and exclusion of SNPs with Hardy-Weinberg ($P<1x10^{-6}$) and with minor allele frequencies (MAF) <1%. Imputation to HapMap variants was performed by Dr So-Youn Shin in Dr Nicole Soranzo's group at the Wellcome Trust Sanger Institute. Briefly, the sparser HumanHap300 dataset was imputed to the HumanHap610Q using phased TwinsUK HumanHap610Q haplotypes as a reference. Next, the combined panel was imputed using reference haplotypes from the HapMap2 project (rel 22, combined CEU+YRI+ASN panels). I obtained this cleaned data then applied further QC, a MAF criteria of 0.05 to ensure rare variants were excluded, thus reducing potential false positives. Of the 960 individuals with microbial 16S data, 886 were genotyped.

2.1.9 *Metabolomics Data*

Serum and plasma metabolite profiles in the TwinsUK cohort were available for 6055 individuals, of whom 831 had available gut microbiome data and were considered for this work. The metabolite data were generated by Metabolon Inc and underwent quality con-

trol (QC) checks and normalisation procedures by Idil Erte, a PhD student within the department. Metabolites were isolated from serum and plasma, and then analysed using ultra-performance liquid chromatography tandem mass spectrometry (UPLC-MS/MS) and gas chromatography mass spectrometry (GC-MS). With the former technique, the sample was loaded into a column with a mobile phase, in this case filled with water, 95% methanol containing 0.1% formic acid or 6.5 mM ammonium bicarbonate, and a stationary phase. This eluent was then subjected to electro-spray ionisation (ESI), sorted by the mass of its ions by electromagnetic fields and detected. The stationary phase for GC-MS was a 5% phenyldimethyl silicone column. The temperature was increased from 60 °C to 340 °C over a 17 minute period before entering the mass spectrometer. Here the sample underwent electron ionisation, then subjected to the mass spectrometer procedures of acceleration, deflection and detection.

These mass spectrometry values were then compared to a reference library that used molecular weight, retention time and in-source fragment values in order to identify the metabolites as known chemical compounds. The metabolite levels were quantile-normalised and any data points that were greater than 4 standard deviations away from the mean were removed. Missing data were imputed using the R package 'mice' [van Buuren and Groothuis-Oudshoorn, 2011] and the quantile-normalized values were further adjusted using linear regression for age, serum/plasma, batch, day and BMI.

Faecal metabolites were determined using UPLC-MS/MS with positive ion mode electrospray, UPLC-MS/MS with negative ion mode electrospray, GC-MS and LC polar platform at Metabolon Inc. Metabolon normalised the values in terms of raw area counts. I then performed further QC on these measurements. I removed any metabolites present in less than 25% of samples as well as any metabolites that was mea-

sured as zero or a constant throughout the entire dataset. Metabolites were inverse normalised then imputed using the R package 'mice' [van Buuren and Groothuis-Oudshoorn, 2011].

2.2 STATISTICAL METHODS

2.2.1 *Linear Mixed Effects Regression*

A significant amount of work contained within this thesis focused on phenotypic associations with 16S profiles in the human gut. All phenotypic analyses were conducted using the statistical program, R [R Core Team, 2014] and the package LME4 [Bates et al., 2014]. The main gut microbiome-phenotype analyses were performed using linear mixed effects regression (LMER), taking into account twin structure. Mixed-effects models are particularly effective at describing the relationship between a response variable, predictors and covariates, which can be fixed effects or random effects. If a set of possible levels is fixed within a variable, and no further observations will yield more levels, for example, no matter the number of extra observations, gender will only have 2 levels: male and female, and as such this covariate would be a fixed effect. However if the observed levels were only a sample of all the possible levels, this would be a random effect. To test for association between OTUs and phenotypes, I compared the fit of two LMER models: a null model which excluded the phenotype/measure of interest, and a full model, which included phenotype/measure of interest in the predictors. All OTUs have undergone prior adjustment for the following technical covariates, number of sequences, technician, sequencing run and shipment batch as well as age. These models are shown in their simplest form below:

$$null : phenotype = sex + (1|famID) + (1|zygosity)$$

$$full : phenotype = OTUs + sex + (1|famID) + (1|zygosity)$$

Using the R package, LME4, I performed the LMER using the models shown above. In both the null and full models, family ID and zygosity were taken as random factors, while in the full model each OTU was added as a fixed effect. Dependent upon the phenotype studied, additional covariates may have been added to the model. Each phenotype was normalised to the Gaussian distribution prior to the analyses, to ensure that each phenotype was normally distributed. The results from this regression were then analysed using an ANOVA (analysis of variance) to determine the significance of the difference between the two models. To account for multiple testing over hundreds of OTUs and taxonomy levels where appropriate, I applied Bonferroni correction, which is overly stringent as OTUs are not independent of each other. In some analyses I also applied Benjamini-Hochberg FDR correction using the R package 'fdrtool' [Klaus and Strimmer., 2015], a more relaxed multiple testing correction. This method takes sorted p-values and divides each p-value by its percentile rank.

2.2.2 *Discordant MZ analyses*

In order to investigate potential non-genetic effects in the microbiome-phenotype results I explored the data using only monozygotic (MZ) twins. Taking only MZ twin pairs, I compared the difference in phenotype values and the difference in OTU values. That is to say, I took the difference between twin 1 and 2 for both the phenotypes and the

OTUs, ensuring always to take the low value from the high to preserve signed differences, and compared these by Pearson correlation in R.

2.2.3 *Twin-Based Heritability Estimates*

Heritability in this dataset is described in detail in Chapter 3.

2.2.4 *Candidate Gene Analysis*

Genetic association testing of human genetic variants with gut microbes was performed at candidate human genes using the open source program GEMMA (Genome-wide Efficient Mixed Model Association, [Zhou and Stephens, 2012]). Firstly, a kinship matrix is calculated to account for familial structure within the data. In these analyses, I calculated the standardized matrix, however a previous study has shown that it makes little difference which method is used to create the kinship matrix [Eu-Ahsunthornwattana et al., 2014]. To perform the candidate genetic association test, I used the univariate linear mixed model which takes a matrix of covariates and considers each phenotype of interest (in this case, OTU) individually. GEMMA tests for association between each specified human SNP and the phenotype of interest, and evidence for association is estimated using 3 test statistics: the Wald test, the likelihood ratio test, and the score test. Results were considered for further investigation if they passed a multiple testing threshold as calculated below:

$$\mathrm{threshold} = \frac{0.05}{\mathrm{n^o loci} \times \mathrm{n^o phenotypes}}$$

where the number of loci refers to the total number of genes tested, and the number of phenotypes refers to the total number of OTUs (and where appropriate taxonomy levels) tested.

3

HERITABILITY OF THE GUT MICROBIOME

3.1 BACKGROUND

Quantitative genetics is concerned with the study of continuous, or quantitative, phenotypes and their underlying genetics. These traits are often influenced by multiple genes but can also be affected by the environment, such as diet, living conditions and even education. Therefore, the variance of a phenotype can be described below:

$$V_P = V_G + V_E$$

V_P = Phenotypic variation

V_G = Genotypic variation

V_E = Environmental variation

V_G can be further subdivided into dominance genetic variance (V_D) and additive genetic variance (V_A) and as such can be represented in the following way:

$$V_G = V_A + V_D$$

Additive genetic variation can be described as the cumulative effect of alleles at a given loci i.e. if AA is white and aa is red, Aa will be pink because the effect is additive. In contrast dominance

genetic variance is described as the interaction between alleles and thus in the same scenario, if AA is white and aa is red then Aa will be white (or red, depending on which allele is dominant), but importantly will not be the average value between the two homozygotes.

Using this framework we can then calculate heritability of a trait in a given population, or more simply, how much of the variance of a trait is attributed to genetics. There are two concepts of heritability that can be calculated; Broad-sense heritability, or H^2, which is the ratio of total genetic variance to phenotypic variance:

$$H^2 = V_G/V_P$$

and narrow-sense heritability, or h^2, which is the sum of additive genetic variance to phenotypic variance:

$$h^2 = V_A + V_P$$

These calculations are not without assumptions however. They ignore the potential interaction between genetics and environment as well as potential epistatic interaction, ie. gene by gene interactions.

Family-based studies are ideal for determining heritability of a given trait due to the relatedness between individuals. There are three potential family-based studies that can be employed for this task; Family studies, adoption studies and twin studies. Family studies are important to establish disease risk for relatives for a disease that another family member may have. Adoption studies are useful for teasing apart environmental differences vs genetic by looking at traits associated with either the biological parents or adoptive parents. Twin studies are perhaps the most popular choice of family study,

particularly for heritability. Using the assumptions on the relatedness between identical (monozygotic) and non-identical (dizygotic) twins, heritability can be calculated for a given trait. It is on twin studies I will focus next.

3.1.1 *Twin Studies and Heritability*

Twin studies provide a simple model for understanding underlying genetic causes due to the ability to compare monozygotic and dizygotic twins. Twins are the perfect controls for each other; they are the same age and they possess very similar early-life environmental exposures. This is in addition to their genetic relatedness. Monozygotic twins are almost 100% identical while dizygotic twins share approximately half their genetic variation. The main assumption of twin studies is that if a trait were caused by host genetics then monozygotic twins would have more similar trait values compared to DZ twins because MZ twins share nearly identical genomes while DZ twins on average share only 50% of their genetic variants, although this is not necessarily correct.

The ACE model is one approach to calculating twin-based estimates of heritability [Neale and Cardon, 1992]. In this approach the classical twin study typically decomposes the phenotype variance into three components that are attributed to genetics (A), common environment (C) and unique environment (E) effects. A number of methods are available to calculate heritability including intra-class correlations (ICCs) and structural equation modelling (SEM).

If we assume that monozygotic twins share 100% of their genetic variants then:

$$rMZ = A + C$$

where, rMZ represents the correlation in trait values in MZ twins. Dizygotic twins however, share half their genetic variation and, we assume, 100% of their common environment therefore:

$$rDZ = 0.5 * A + C$$

where rDZ is the correlation in trait values in DZ twins. Using this assumption, A, C and E can be estimated no-matter which method of calculation is used.

3.1.2 *ICC Method*

The intraclass correlation (ICC) is a statistical method used to describe how similar two groups are to each other. This method can be used to calculate heritability by calculating the ICC between twins, resulting in rMZ and rDZ. The A component can be estimated by taking twice the difference between ICCMZ and ICCDZ as monozygotic twins are twice as similar as dizygotic twins:

$$A = 2(ICCMZ - ICCDZ)$$

As we know the ICC is the sum of A and C, we can estimate C by taking A from rMZ:

$$C = ICC - A$$

Finally, to calculate E (unique environment) we consider that A + C + E is equal to 1, and so we can subtract ICC from 1:

$$E = 1 - ICC$$

3.1.3 *Structural Equation Modelling and OpenMX*

Structural equation modelling (SEM) is an alternative approach to estimate twin-based heritability, based on path analysis and regression. OpenMX is a package in R that conducts SEM to estimate heritability of a given trait. OpenMX provides two methods of modelling that can be employed; path analysis and matrix specification. Path analysis uses a path diagram in order to calculate the latent variables, in the case of heritability, A, C and E. Figure 1 shows the path diagram for twin ACE heritability modelling in OpenMX.

The second method of SEM modelling employed by OpenMX, and used to calculated heritability in this thesis, is matrix specification. Matrices are specified for the path coeefficients a, c and e and then to calculate variance matrices for A, C and E are calculated by squaring the path coefficients. MZ and DZ covariance matrices are calculated according to the assumptions shown in Figure 1. Selection of a particular method comes down to the preference of the user.

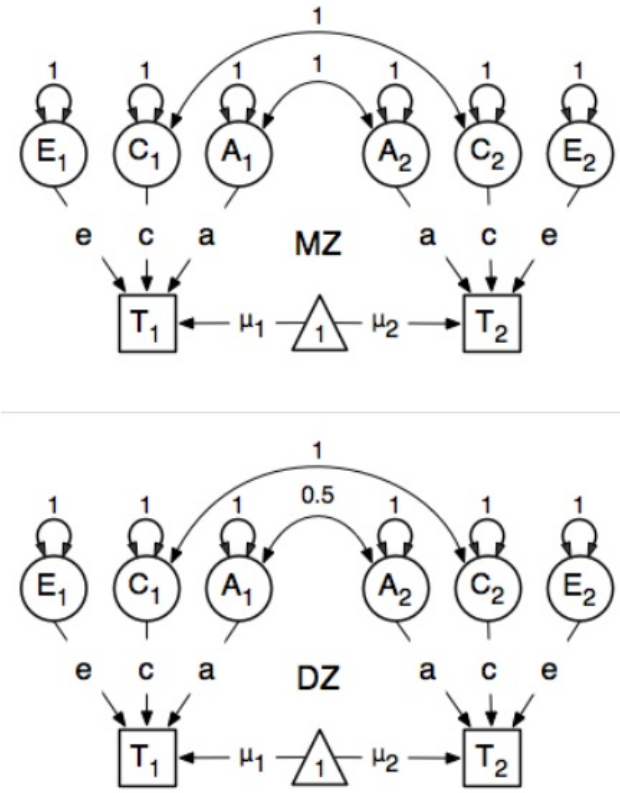

Figure 1: Path diagram for the calculation of heritability in monozygotic and dizygotic twins.(Image taken from OpenMX user manual: http://openmx.psyc.virginia.edu/docs/OpenMx/latest/OpenMxUserGuide.pdf)

3.1.4 *Evidence Of Genetic Effect On The Microbiome*

Many studies have attempted to understand whether there is an underlying host genetic component influencing the microbiome, but a clear explanation has not yet been found. A number of approaches have been used to address the question of heritability such as twin studies, mouse models and QTL detection.

Two 16S genetic studies into the human gastrointestinal microbiota have been performed using twins [Zoetendal et al., 2001, Stewart et al., 2005]. In both cases the results were suggestive of a host genetic basis for the gut microbiota however neither of these studies mention which aspects of the microbiota are heritable, only that the

similarity in faecal bacteria between monozygotic twins was significantly higher than dizygotic twins. Alternatively, a different twin study using 16S rRNA and metagenomics found no basis for heritability [Turnbaugh et al., 2009a]. However, components of diet and specific lifestyle choices have previously been shown to be heritable [Teucher et al., 2007, Chaput et al., 2014] and these may have a great impact on the microbial species present within the gut.

Mouse models have been useful for a wide range of studies including gene knock-out models where certain genes are knocked out and the effect this has on a given phenotype is observed. Studies have demonstrated that the microbiota of mice in the same litter is more similar than mice from a different litter, however this is may be a maternal effect, whereby mice are inoculated by the mother's microbiota while being born vaginally. There are approaches to try and reduce this effect in order to study genetic influences such as cross-fostering and uterine transplants of embryos in genotypically different mothers. The latter approach was used by Friswell et al. [2010] but rather than discovering a genetic influence, they found a second maternal effect.

Using large numbers of intercross model animals it is possible to explore both the impact of host genetic background and environmental factors. One of the first studies to use this approach in order to characterise genetic influences on the microbiome was that by Benson et al. [2010]. Among other things, they found the presence of a core microbiome composed of 64 taxonomic groups that were present in most, if not all of the mice sampled. It could be suggested that the presence of such a conserved group in all mice would indicate a genetic relationship, however as yet no definitive genetic link has been found. These findings confirm the results found in Qin et al (2010) where 35% of reads from any one sample could be mapped

to other samples thereby hinting at the existence of a common core microbiome [Qin et al., 2010].

There is evidence for the effect of individual genes on the microbiome in both humans and mice. Mice lacking the *RELMB* gene have significantly altered abundances of Bacteroidetes, Firmicutes and Proteobacteria compared with wild-type mice [Hildebrandt et al., 2009]. The mechanism underlying this finding is unknown but it was noted that mice lacking this gene did not become obese when fed a high-fat diet. A number of the genetic variants, that have so far been linked with the microbiome, cause disease within the host, for example a single nucleotide polymorphism (SNP) within the human *MEFV* locus which encodes pyrin, a key protein in white blood cells that is thought to help regulate inflammation, results in familial Mediterranean fever whilst also being associated with lower microbial diversity [Khachatryan et al., 2008]. Further interactions have been noted for Crohn's disease [Frank et al., 2011], Coeliac's disease [Vaahtovuo et al., 2003] and obesity [Zhang et al., 2010].

A problem with these studies is small sample size. In order to detect host genetic effects, that often have a small effect size, a sample size of thousands of subjects is required, even hundred's of thousands as we have seen with genome-wide association studies of complex traits.

In this chapter I will address heritability of the human gut microbiome in twins and explore a number of potential environmental factors, including diet, method of delivery and smoking, that may affect the host gastrointestinal flora.

3.2 RESULTS AND DISCUSSION

3.2.1 *Heritability Of The Human Gut Microbiome*

Our study, a collaboration led by Dr Ley (Cornell University), was the first large-scale study to characterise heritability for all gut microbe OTUs in a dataset of 416 twin pairs [Goodrich et al., 2014]. The dataset was comprised of 342 monozygotic twins, 490 dizygotic twins and 143 singletons, predominantly female. The aim of the work was to characterise heritable microbes in the human gut and if there were heritable microbes, determine how they were related to BMI. Using OpenMX, Ms Goodrich (Cornell University) calculated heritability for each OTU and collapsed taxonomy.

While my colleagues at Cornell used OpenMX to calculate narrow-sense heritability, I used the ICC method of estimating heritability to confirm their results. I considered 909 microbial units, 768 OTUs and 141 taxonomic summaries.

Figure 2 shows the correlation between my heritability estimate from ICC, and the heritability estimated from Cornell university using SEM. The Pearson correlation coefficient for A and E were 0.55 and 0.54 respectively. The C component however only had a correlation coefficient of 0.3. A table of all the microbial units and their ICC and ACE heritabilities can be found in Appendix A. The peak ICC heritable microbe was a Ruminococcaceae OTU (2*(rMZ-rDZ) = 0.72). ICC heritability ranged from 0-0.72.

We also showed that microbiota diversity is more similar between twins than unrelated individuals using weighted and un-weighted Unifrac distances, as well as the Bray Curtis metric (Student's T-Test, $p < 0.009$). Furthermore, monozygotic twins have more similar microbiotas than dizygotic twins, but only when the un-weighted

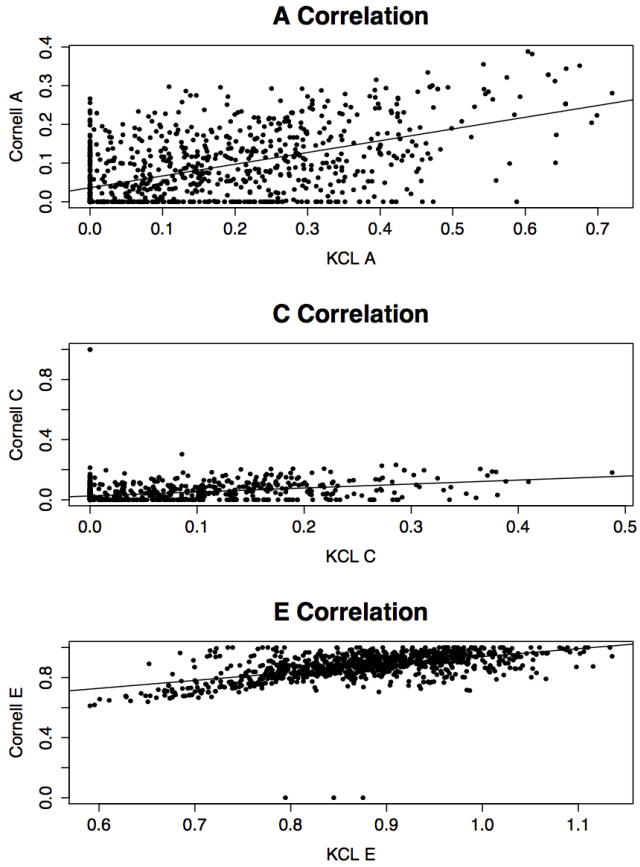

Figure 2: Correlation of heritability components calculated by ICC and SEM.

Unifrac metric was considered (p = 0.032), suggesting that types of species shared drive similarities between monozygotic twins rather than abundance similarity, as weighted Unifrac takes OTU abundance into account when calculating Unifrac distances.

The heritability estimates from SEM in this dataset (calculated by Ms Goodrich) are shown in Figure 3 below at each node of the phylogenetic tree. Certain branches of the families Ruminococcaceae and Lachnospiraceae were found to be highly heritable, with heritability ranging from 0-0.38 (Figure 3). A complete table of both ACE and ICC heritability estimates can be found in Appendix A (Table A1).

These heritabilities were then replicated in 2 independent datasets Turnbaugh et al. [2009a] & Yatsunenko et al. [2012]). The

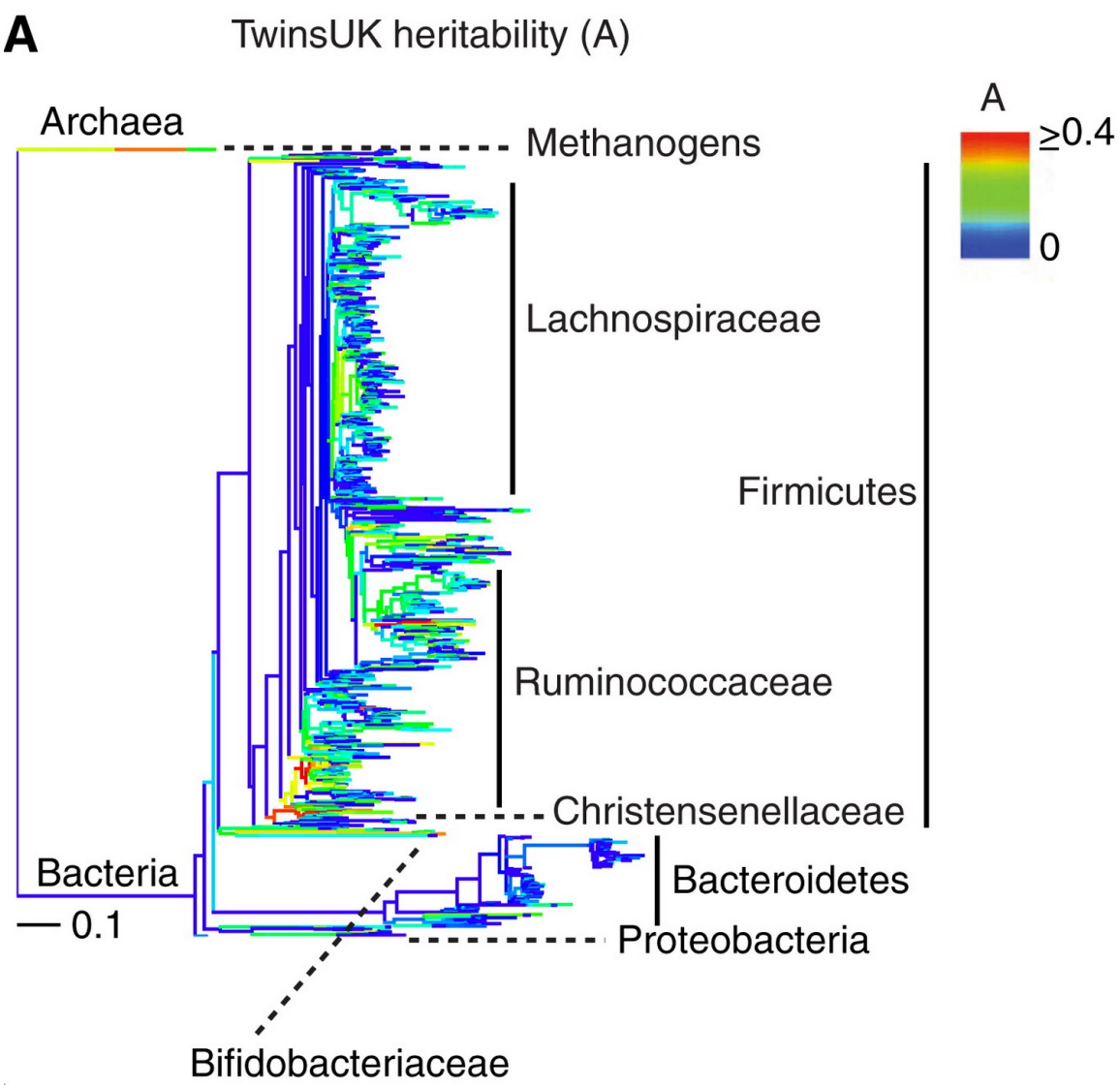

Figure 3: Heritability of OTUs in the TwinsUK cohort. Reproduced from Goodrich et al, 2014 [ref:goodrich2014]

most heritable microbe was *Christensenella*, a microbe that was characterised relatively recently and as such little is known about it. Ms Goodrich created co-occurence networks and found that Christensenellaceae formed the centre of a highly heritable hub of microbes, including the interesting Archaea member, Methanobacteriaceae. Methanobacteriaceae are capable of producing methane naturally and are key organisms in the fermentation of complex compounds in the gut [Pimentel et al., 2012]. Microbes in this co-occurrence hub were also enriched in lean individuals, and may play a key role in obesity,

its prevention or downstream effects on the human organism. When *Christensenella* was spiked into mice the association with obesity was confirmed, as mice that possessed *Christensenella* showed less weight-gain than those that had no *Christensenella*. That a highly heritable microbe has such an effect on weight-gain suggests the importance for host genetics in microbiota interactions with health.

3.2.2 *Environmental Effects On The Microbiome*

In the ACE results from SEM and ICC we observed that for most OTUs, the unique environment component (E) accounted for the highest proportion of variance in the OTU showing the importance of environmental effects on the microbiome. A number of environmental effects on the microbiome have been studied, such as method of birth and diet, with the latter having been studied most extensively.

3.2.2.1 *Diet*

Arguably, one of the most important factors affecting the human gut microbiome is diet. The microbes that live in our guts depend on the food we eat for their own sustenance, while helping to digest our food in the process. For this reason, small shifts in what we eat can change microbial community composition significantly. Typically, in large cohort-based studies such as this, diet information is collected from study participants using a questionnaire known as a Food Frequency Questionnaire (FFQ). Participants are asked to estimate how often they eat a particular item of food and indicate an approximate portion size that they would normally consume. From this, nutrients can be derived and food groups created based on the similarity in nutrient content between items.

In a study of elderly subjects (ages between 64-102) of Irish nationality it was concluded that a healthy, diverse diet correlated with a more diverse gut microbiome [Claesson et al., 2012]. Using FFQ data, they found four dietary groups: DG1 (low fat/high fibre), DG2 (moderate fat/high fibre), DG3 (moderate fat/low fibre) and DG4 (high fat/low fibre). DG1 showed the highest microbial and dietary diversity and as expected, DG3 and DG4 showed the lowest diversity. The study was unable conclude whether diversity of the diet affects microbial diversity, or if the ratio of fat:fibre was key.

Metagenomic analysis of gnotobiotic mice that had been transplanted with stool samples from adult humans and placed on either a Western diet or a low fat, plant polysaccharide-rich diet (LF/PP) revealed a number of species that differed between the two diets [Turnbaugh and Gordon, 2009]. Western diet humanised mice had higher levels of *Erysipelotrichi* species when compared to mice on the LF/PP diet. In addition, there were significant increases in the relative abundances of *Bacilli* and decrease in Bacteroidetes. Adiposity gains were seen in mice that were fed the Western diet, but not in those fed the LF/PP diet. When new mice were transplanted with cecal contents from these humanised mice, those that had a Western diet donor gained significantly more fat than those mice with LF/PP donors.

Both of these studies have focused on high fat/low fat diets and implications with obesity but a recent study into use of artificial sweeteners sheds a new unusual light on the role of the microbiome in obesity. Artificial sweeteners are often used in diet soft drinks and sugar-free sweets, purporting to be the healthy alternative to other high calorific products. In this study, mice that had been fed saccharin showed an increase in glucose intolerance both in the lean state ($p < 0.001$) and obese state ($p < 0.03$) [Suez et al., 2014]. When antibi-

otics were given to mice consuming saccharin, glucose intolerance disappeared and mice were no different than the controls, indicating a role for the microbiota in driving the glucose intolerance. Mice consuming saccharin displayed a significant dybiosis from controls including increases in relative abundance of taxa within the *Bacteroides* and Clostridiales. Furthermore, these mice had altered bacterial functional profiles, showing marked increases in glycan degradation resulting in high-energy harvest. The study also extended its findings regarding glucose intolerance to humans volunteers using diet questionnaires and blood measures of glucose. As glucose intolerance is seen in a high number of obesity cases, this study has important implications for the role of sweeteners in cardiometabolic disease and the growing obesity epidemic and has provided a sound basis for the role of the microbiome in propagating this glucose intolerance.

As there is a wealth of evidence supporting the change that diet can make in the gut microbiome, I used FFQ diet data for 838 individuals that had been derived in a previous study on twins (see Chapter 2 for further information.) [Teucher et al., 2007] to determine whether diet associates with changes in the gut microbiome in the TwinsUK cohort. To begin with, I looked at the first 5 principle components that roughly equated to 5 dietary patterns: PC 1 was a fruit and vegetable diet, PC 2 was a high alcohol diet, PC 3 was the traditional English diet, characterised as a diet high in fried fish, pies and meat as well as some cruciferous vegetables. PC 4 was a dieting profile, high in low-calorie drinks and low-fat dairy products. The final score, PC 5 was a low meat diet. As expected, each score was normally distributed. A linear mixed effects regression was used to test the association between each diet PC and OTU, adjusting for zygosity and family structure. In total, 4545 tests were performed. The results can be found in Table 4.

OTU	Family	Genus & Species	Diet	P Value	Std Err	Estimate	FDR
759816	Ruminococcaceae	Unknown	High Alcohol	2.86E-07	0.0644	0.3334	0.00114
232828	Ruminococcaceae	Unknown	High Alcohol	5.00E-07	0.0547	0.2772	0.00114
Collapsed	Ruminococcaceae	Ruminococcus	Fruit and Veg	5.89E-06	0.0047	-0.0216	0.00702
Collapsed	RF32 (Order)		Low Meat	9.27E-06	0.0230	-0.1026	0.00702
Collapsed	RF32 (Unknown)		Low Meat	9.27E-06	0.0230	-0.1026	0.00702
Collapsed	RF32 (Unknown)	Unknown	Low Meat	9.27E-06	0.0230	-0.1026	0.00702
Collapsed	Odoribacteraceae	Butyricimonas	High Alcohol	5.08E-05	0.0130	0.0531	0.03114
179744	Clostridiales (Order)	Unknown	Fruit and Veg	8.25E-05	0.0706	0.2798	0.03850
Collapsed	Lachnospiraceae	Lachnospira	Traditional English	8.63E-05	0.0109	-0.0432	0.03850
190961	Lachnospiraceae	Unknown	High Alcohol	9.32E-05	0.0448	0.1761	0.03850
194488	Ruminococcaceae	Unknown	Traditional English	0.000103	0.0575	-0.2255	0.03903
258099	Ruminococcaceae	Unknown	Fruit and Veg	0.000119	0.0645	-0.2495	0.04146
2318497	Clostridiaceae	Unknown	Fruit and Veg	0.000152	0.0232	0.0883	0.04761
4434334	Clostridiaceae	Unknown	Low Meat	0.000167	0.0114	0.0431	0.04761

Table 4: Significant diet associations with the human gut microbiome that pass an FDR correction threshold of 5%.

A multiple testing threshold was applied to the results using Bonferroni correction, estimated at P = $1.1x10^{-5}$ and 6 associations surpass this. 14 associations pass an FDR threshold of 5% and 485 associations were nominally significant, passing a p-value threshold of 0.05. The top association was between OTU 759816 (Ruminococcaceae) and PC2 (High Alcohol) with a p-value of 2.86E-07. In fact the High Alcohol diet PC held a large number of associations with OTUs in not just the Ruminococcaceae family, but also other families such as the Odoribacteriaceae, (*Butyricimonas*, p = 5.08E-05, q = 0.0311), or in the Pasteurellaceae family, in particular *Haemophilus parainfluenzae* (p = 0.00075, q = 0.11). In some cases, these associations are positive suggesting that an increase in OTU abundance coincides with increased alcohol consumption. It is unclear in which direction this relationship goes without an in-depth look into causal models, but one potential explanation for these associations is that these microbes aid in the metabolism of alcohol.

PC 1, a fruit and veg diet, also had a number of nominally significant associations, most notably with the *Ruminococcus* collapsed taxonomy (p = 5.89E-06, q = 0.007) and OTU 179744 (Clostridiales; Unknown species, p = 8.25E-05, q = 0.038). Fruit and veg is a significant source of fibre and many studies have shown how fibre affects the

microbiome, promoting a more diverse microflora as well as widely regarded beneficial species such as *Faecalibacterium prausnitzii*.

While it is interesting to look at the diet principle components, they do not capture exact diets. For example, the fruit and veg diet is characterised as a diet that is high in fruit and vegetables, but people could also be eating a large amount of fatty foods alongside fruit and veg. A more accurate way to assess the effect of diet on the microbiome is to look at nutrient intakes, such as the breakdown of different fats. It was interesting to note that despite the large wealth of evidence in studies to date showing the effect of dietary fat on the microbiome, we saw few significant associations between OTUs and the traditional English PC, which we considered to be that most alike to a high fat diet. So instead we looked at dietary fat intakes to determine whether the PC wasn't an accurate proxy for high fat intakes or whether fat intakes had little effect on this sample. In this part of the analysis we considered the following intakes: trans fat, saturated fat, monounsaturated fat, polyunsaturated fat and total fat, all of which were adjusted for energy intake. My colleague Ms Tess Pallister extracted the nutrient variables from available FFQ data and I performed an association analysis between these fat intakes and gut 16S data. The top association between trans fat and the collapsed taxonomy for *Ruminococcus* was the only association that surpassed FDR correction for multiple testing (p = 2.70E-07, q = 0.001). There were a further 383 nominally significant results.

3.2.2.2 *Method Of Birth*

Until recently it was thought that a baby did not come into contact with microbes until after birth. The womb was considered to be a sterile environment except in cases of disease, but recent findings have changed this view. Amniotic fluid, a litre of which is swallowed by

a foetus each day, is teeming with microbial life, giving the foetus its first inoculation of bacteria [DiGiulio, 2011, Jiménez et al., 2005]. A new study has shown that even the placenta has a microbiota, giving further evidence to abolish the idea of a sterile womb [Aagaard et al., 2014]. The largest transmission of microbes to the foetus however is during birth [Dominguez-Bello et al., 2010]. The vagina is a microbially-rich site mainly dominated by lactic-acid bacteria that help to maintain a constant low pH and play a key role in protecting against urogenital illness [Ravel et al., 2011]. If a child is born vaginally, the foetus passes through the birth canal and is bathed in these bacteria, but if born by caesarean section, it misses out on this dose of microbes. A study by Dominguez-Bello et al. [2010] revealed that children born vaginally have skin microbes similar to their mother's vagina and can be assigned to their mother based on these profiles. Children that had been born by C-section however were no more alike to their mother than any other person, and their microbiome was more alike to a skin microbiota, showing enrichment of pathogenic species such as *Staphylococcus*. This may have significant implications for newborn health, for example, the incidence of methicillin-resistant *Staphylococcus aureus* infection (MRSA) in C-section babies is much higher than vaginally-born babies, perhaps because of the higher presence of Staphylococcus species. Of course, this is not a causal link and other factors may be affecting this observed effect, such as the prevalence of premature births in C-section deliveries, but it does pose an intriguing insight into potential health issues posed by method of delivery.

Contraction of MRSA infection is not the only illness thought to be influenced by method of delivery. There is evidence to suggest that obesity [Huh et al., 2012], asthma [Thavagnanam et al., 2008] and allergy [Negele et al., 2004, Eggesbo et al., 2003] are all affected

by whether or not a child was born vaginally. In a follow-up study however, Dominquez-Bello and colleagues showed that the a normal birth microbiota can be partially restored to caesarean-born infants by exposure to maternal vaginal fluids at birth [Dominguez-Bello et al., 2016]. If a delivery method induced microbial dysbiosis is responsible, at least partially, for these illnesses, then this study provide promising results for future treatments and therapies.

The TwinsUK cohort is well phenotyped and information regarding method of delivery was available for a subset of 536 individuals (see Chapter 2 for complete sample characteristics.). In the UK, caesarian section is currently the method of birth for approximately 60% of twin births, however in the 1950's, around the time a large proportion of our sample was born, the incidence rate was 3%. In the sample presented within this thesis, caesarean section accounted for 3.9% of the sample, being representative of the time. I wished to explore whether method of delivery effects on the human gut microbiome might still be present in later life. Using a similar linear mixed effects model as outlined in Chapter 2, we placed the OTU as the dependent variable within the model and considered method of birth as a fixed effect. As expected, the results were not particularly significant, with only 22 associations passing a nominal p-value threshold of 0.05 (Table 5).

The top association with OTU 180352 (Clostridiales; unknown) had a p-value of 0.007 and a negative direction of effect, potentially suggesting a reduction in abundance of this OTU after a C-section, but not passing a multiple testing threshold. Considering there are such low numbers of C-section twins in this sample, there is not enough power to determine differences in the microbiota, especially so much later in life.

OTU	Family	Genus & Species	P Value	Std Err	Estimate
180352	Lachnospiraceae	Unknown	0.0073	0.1608	-0.4381
176865	Ruminococcaceae		0.0078	0.1376	-0.3675
182874	Bacteroidaceae	Bacteroides (Unknown)	0.0138	0.2135	0.5307
4395096	Bacteroidaceae	Bacteroides ovatus	0.0138	0.1132	0.2810
340474	Bacteroidaceae	Bacteroides (Unknown)	0.0190	0.1110	-0.2626
578016	Porphyromonadaceae	Parabacteroides distasonis	0.0201	0.3218	-0.7503
216710	Ruminococcaceae	Unknown	0.0203	0.3303	0.7685
183970	Lachnospiraceae	Blautia (Unknown)	0.0216	0.1669	-0.3861
182255	Bacteroidaceae	Bacteroides (Unknown)	0.0268	0.1531	0.3408
175836	Ruminococcaceae	Unknown	0.0275	0.0896	-0.1989
186416	Lachnospiraceae	Unknown	0.0306	0.6308	1.3670
2875735	Bacteroidaceae	Bacteroides (Unknown)	0.0332	0.1135	0.2425
4381553	Bacteroidaceae	Bacteroides (Unknown)	0.0340	0.6264	-1.3344
179847	Lachnospiraceae	Blautia (Unknown)	0.0360	0.1042	-0.2197
3439403	Bacteroidaceae	Bacteroides (Unknown)	0.0374	0.2175	0.4542
180401	Lachnospiraceae	Blautia	0.0375	0.1161	-0.2426
191779	Lachnospiraceae	Bacteroides (Unknown)	0.0380	0.1133	-0.2356
157453	Ruminococcaceae	Unknown	0.0395	0.3403	0.7049
403701	Veillonellaceae	Dialister (Unknown)	0.0403	0.4066	-0.8371
193551	Lachnospiraceae		0.0429	0.1339	-0.2722
3588390	Bacteroidaceae	Bacteroides (Unknown)	0.0441	0.5989	1.2081
190649	Ruminococcaceae	Unknown	0.0468	0.1014	0.2026

Table 5: Nominally significant associations between the human gut microbiome and method of delivery.

3.2.2.3 *Smoking*

Smoking has become one of the leading causes of preventable death worldwide, killing over 480,000 people each year in the US alone [General, 2014]. In fact, smoking is a risk-factor for not just lung cancer, emphysema and pulmonary disease but also irritable bowel disease (IBD) and ischaemic heart disease along with a much reduced lifespan. Yet despite the adverse health conditions associated with smoking, it remains a popular habit and drives a thriving multi-billion pound industry. Heritability of smoking behaviour is reportedly quite high at 44% and 55% of the variance is driven by shared environment [Vink et al., 2005]. Interestingly, the heritability of nicotine dependence is even higher at 75%, having intriguing implications for people wishing to quit smoking and suggesting that some people may find it much more difficult to quit than others [Vink et al., 2005].

With such inhalation of toxins in the smoke, it is logical to assume that there would be an effect on the microbiome in the lung. Yet new research has come to light showing that smoking also affects the gut microbiome, with a potential influence on the weight gain so often seen after smoking cessation [Biedermann et al., 2013]. Participants in the smoking cessation group had a 2.2kg mean increase in body weight, supporting previous findings [Aubin et al., 2012] and this was not due to differences in calorie or nutrient intake. Furthermore, the relative abundance of Firmicutes and Actinobacteria significantly increased after smoking cessation, while the relative abundance of Proteobacteria decreased after the participant had stopped smoking, similar to that seen in comparisons between lean and obese subjects. Each study group could be clearly separated using Unifrac showing clear differences in beta diversity while alpha diversity was not significantly different between smokers and non-smokers.

In my dataset I had smoking status for a total of 851 people (534 non-smokers, 262 ex-smokers and 55 smokers). Firstly I looked at alpha diversity vs smoking status, as measured by PD Whole Tree Averages. There was no significant difference between smoking status and alpha diversity although a negative trend could be seen with a minor reduction in alpha diversity in smokers (Figure 4).

Next I looked at beta diversity between smokers, ex-smokers and non-smokers. 3 metrics are used to measure beta diversity: weighted and unweighted unifrac metrics and Bray Curtis. No metric showed any clustering according to smoking status (Figure 5) suggesting that smoking does not significantly alter microbial diversity in the gut.

With this in mind, I wanted to see if any individual OTUs changed significantly in smokers. Using a linear mixed effects model, I assigned each OTU to be the response variable and adjusted for family, zygosity and sex. The top association was with OTU 4426298,

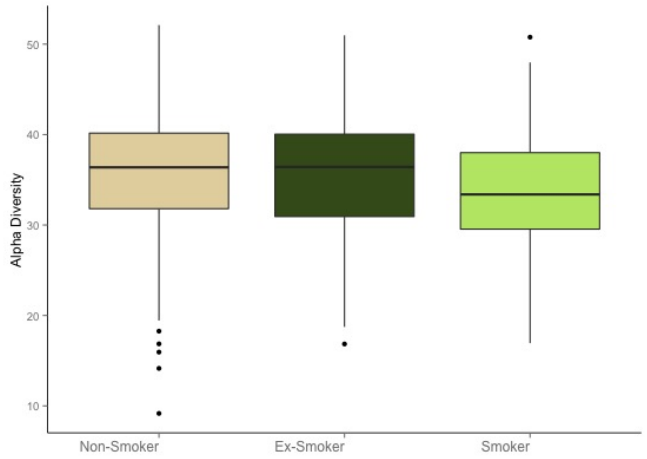

Figure 4: Boxplots of PD Whole Tree Averages for each smoking status group.

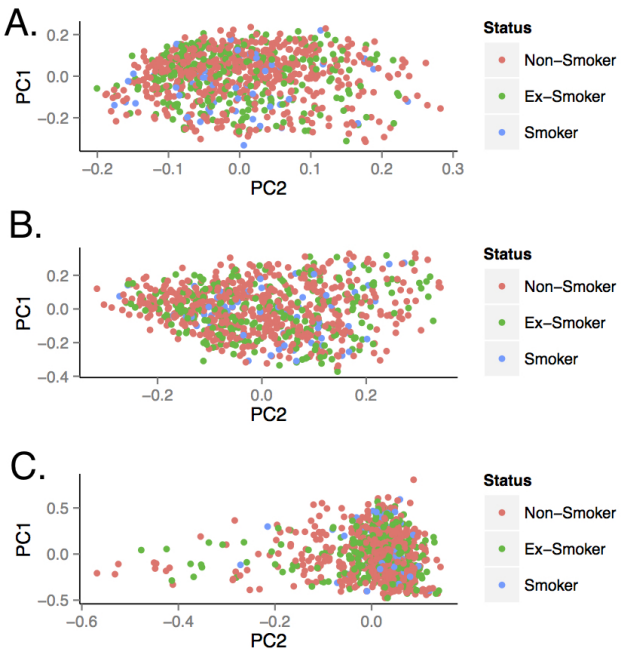

Figure 5: PCoA plots showing diversity in the 3 beta diversity metrics, coloured according to smoking status. A. Unweighted Unifrac metric. B. Bray-Curtis metric. C. Weighted Unifrac metric.

a *Bifidobacterium animalis* OTU (p=0.0005, β=-0.38, SE=0.12 Figure 6). While this result does not pass Bonferonni there is a clear stepwise reduction in *B. animalis* from smokers through to non-smokers. After further adjustment for BMI, the p-value remained nominally significant (p=0.001).

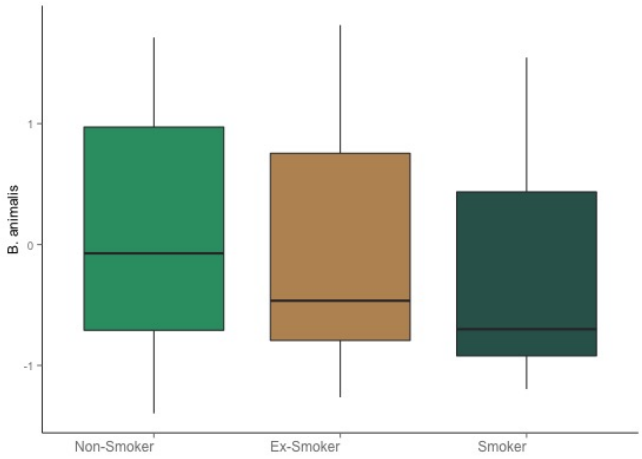

Figure 6: Boxplot of *B. animalis* in smokers, ex-smokers and non-smokers.

B. animalis is a well-known probiotic, used to aid gut health and reduces symptoms of bloating [Philippe et al., 2011, Guyonnet et al., 2007]. Here we see it negatively associated with smoking status showing a marked decrease in smokers. Research in smokers has shown that smoking puts them at increased risk of gut Crohn's disease [Cottone et al., 1994] but the mechanism of this effect is poorly understood. Reduction of *B. animalis* may be part of the reason if this is the case, and provides a potential treatment option for smokers suffering from poor gut health. While overall diversity remains unchanged in smokers, differences at the OTU level may have some effects on health.

3.3 CONCLUSION

In this chapter I have explored the effects of host genetics using heritability analysis and the impact of major environmental factors thought to affect the human gut microbiome in a large sample of adult twins. Some microbes in the human gut microbiome are moderately heritable, but a large role can be attributed to environmental factors. In this cohort there appear to be no strong, long-term effects

on the gut microbiome from method of delivery, while smoking may have a small effect on gut community composition. Diet showed a number of associations with gut microbial composition in our sample, but the findings leave questions unanswered due to the difficult nature of measuring dietary intakes accurately, but there are some effects seen with dietary fat. As *Christensenella* was found to be the most heritable microbe, I was interested to see what else I could discover about this relatively unknown bacteria. Chapter 4 is about my work characterising this microbe further.

4

CHRISTENSENELLA

4.1 BACKGROUND

Christensenella is an obligate anaerobe, like most bacteria in the human gut microbiome. Its characterisation was recent [Morotomi et al., 2012] and as such little is known about this seemingly important microbe. It produces butyrate and as well as being present in human samples, the characterised strain shares 96.7% sequence identity with an uncultured clone found in a captive *Dugong dugon*, commonly known as the dugong [Morotomi et al., 2012, Tsukinowa et al., 2008].

Our collaborative work, led by Dr Ley and lead analyst Ms Goodrich, establishing the heritability of the human microbiome [Goodrich et al., 2014] (Section 3.2.1) led to the discovery that *Christensenella* is the most heritable microbe in our cohort of twins as well as in 2 other datasets (Yatsunenko et al. [2012] and Turnbaugh and Gordon [2009]), and approximately 39% of its variance can be attributed to host genetic influence. It co-occurs with the methanogen-producing Archaea and is more abundant in lean twins. Network analysis revealed that Christensenellaceae forms the centre of a heritable hub of microbial families, including the Dehalobacteriaceae, Methanobacteriaceae, SHA-98 and RF39.

These microbes were also significantly associated with a lean BMI (indicated as a red or orange node, Figure 7B). In order to confirm the methanogen-Christensenellaceae co-occurrence, faecal transplants were performed on germ-free mice, inoculating them with ei-

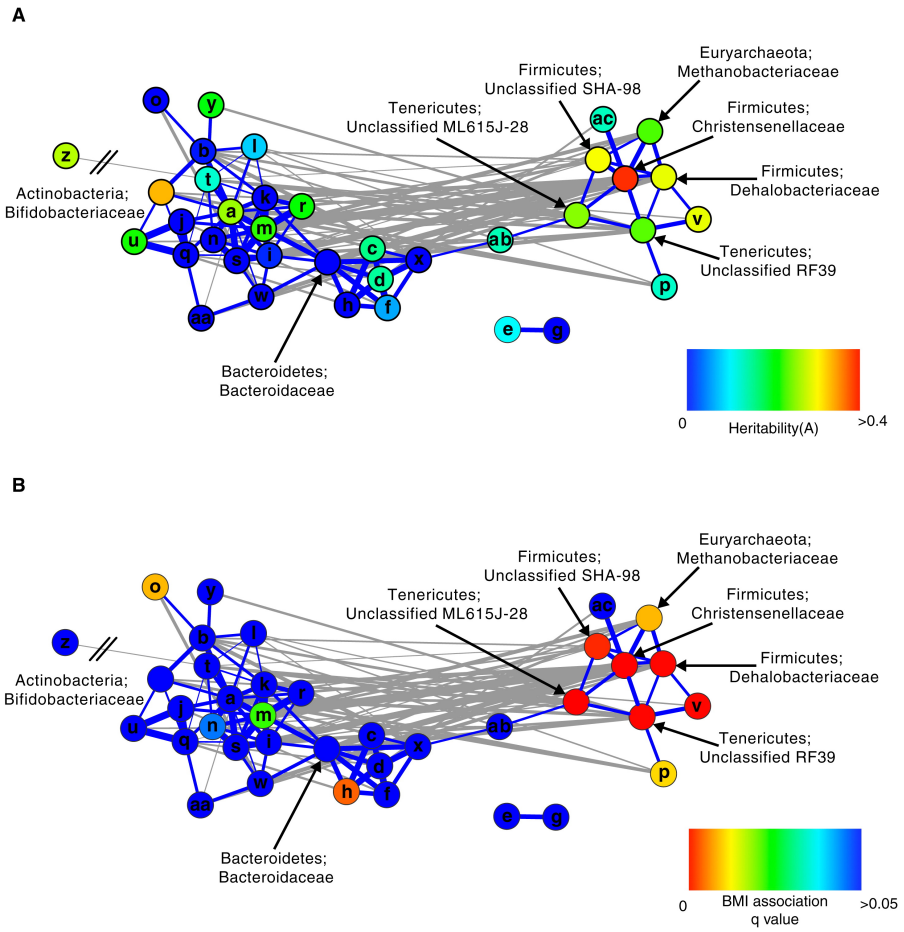

Figure 7: Network of *Christensenella* and connected microbes. A) Network module of OTUs that co-occur with *Christensenella*. Nodes are coloured according to their heritability. Red is heritable while blue is not heritable. B) Network module of OTUs that co-occur with *Christensenella*. Nodes are coloured according to their FDR-adjusted p-value with BMI. Red is highly significant while blue is not significant (P>0.05). Reproduced from Goodrich et al. [2014]

ther a microbiome from an obese donor twin with no methanogens (O-), an obese donor twin with methanogens (O+), a lean donor twin with no methanogens (L-) or a lean donor twin with methanogens (L+). Christensenellaceae is present throughout the study in all but one donor group, and mice that received a transplant from a L+ or O+ donor showed the highest relative abundance of Christensenellaceae throughout the study confirming the co-occurrence of Christensenellaceae with the methanogens (Figure 8A).

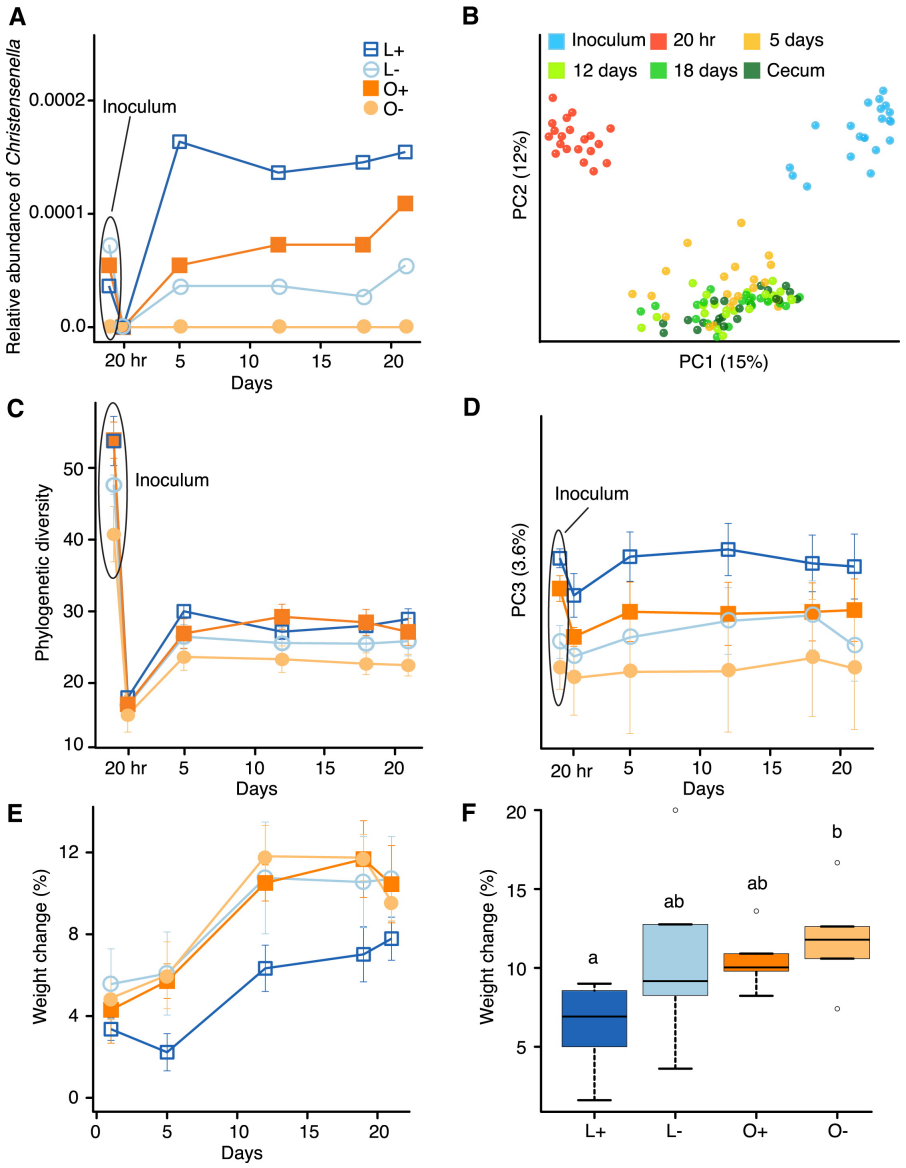

Figure 8: Germ-free faecal transplantation with donor stool from obese and lean twins that either possess or don't possess methanogens. Reproduced from Goodrich et al. [2014]

Furthermore, mice that had received stool from the L+ donor showed reduced weight gain in comparison to mice that had received stool from the other donor groups (Figure 8E,F) as well as possessing the highest alpha diversity (Figure 8C). In order to infer causality of Christensenellaceae, stool from an obese donor that had no Christensenellaceae was spiked with *Christensenella minuta* and transplanted into mice. Mice that received *C. minuta* showed attenuated

weight gain in comparison to mice who did not receive *C. minuta* (Figure 9A).

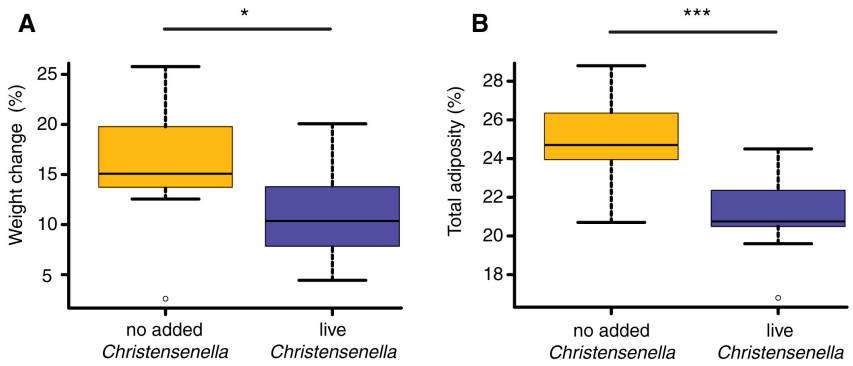

Figure 9: Mice spiked with *C. minuta* show less weight gain than those who did not receive *C. minuta*. Reproduced from Goodrich et al. [2014]

This difference was also seen in the total adiposity in the mice (measured by whole body dual energy x-ray absorptiometry (DEXA) scanning, Figure 9B), a difference that was more significant than % weight change, showing that in particular, *Christensenella* was affecting the adipose tissue of mice. This indicates a potential protective role for Christensenella in the prevention of adiposity gain.

Our previous collaborative study, Goodrich et al. [2014], highlighted the importance of *Christensenella* in obesity, both in twins and in mice. The previous study was limited to BMI however. Here, I extend this work by exploring the association of *Christensenella* with multiple adiposity phenotypes in twins including central adiposity, body fat distribution and the difficult to measure, but clinically relevant measure, visceral fat. As a result of these findings, I proceeded to investigate the association between *Christensenella* and immune-related phenotypes. Lastly, I linked the results to the observation that *Christensenella* is one of the most heritable gut species, by comparing human genentic variation with its relative abundance in the context of immune traits.

4.2 RESULTS

4.2.1 *Christensenella associates with multiple adiposity measures*

Using an LMER model as outlined in Section 2.2.1, I explored the association between the abdominal adiposity measures visceral fat mass (VFM) and % trunk fat (pTF) as well as the whole-body metric, body-mass index (BMI) in 960 individuals (Table 6). Abdominal adiposity, or central adiposity, is the measure of visceral and subcutaneous fat in the abdominal region. Increased central adiposity is also highly correlated with adverse cardiometabolic outcomes [Yusuf et al., 2004]. Visceral fat is the adipose tissue that surrounds organs while subcutaenous fat is found beneath the surface of the skin. It is in fact visceral fat that has the most implications for health. I discuss adiposity in more detail in Chapter 5. Heritability of these measures were also explored and the results of that analysis can be found in Section 5.2.1.

Phenotype	Abbreviation	Description	Sample Size	Time between samples
Visceral Fat	VFM	Deep adipose tissue that forms around organs, composed of adipocytes.		
% Trunk Fat	pTF	% of central mass that is formed of adipose tissue (both visceral and subcutaenous).	960	0-16 years
BMI	BMI	Body mass distribution.		

Table 6: Phenotypes used in adiposity *Christensenella* association study.

At both the family level (Christensenellaceae) and a genus level (Christensenellaceae.unknown genus), there were a total of 17 nominally significant associations (Table 7). The peak association was between an unknown genus in the Christensenellaceae family and VFM ($P=3.51x10^{-6}$, β=-0.0071, SE=0.0015). Both VFM and pTF (peak pTF association $P=2.21x10^{-5}$, β=-0.0542, SE=0.0127) formed stronger associations with Christensenellaceae than BMI (peak BMI association P=0.0008, β=-0.0429, SE=0.0126).

OTU	Family	Genus & Species	Pheno	P Value	Std Err	Estimate
Collapsed	Christensenellaceae	Unknown	VFM	3.51E-06	0.0015	-0.0071
Collapsed	Christensenellaceae		VFM	5.03E-06	0.0015	-0.0071
Collapsed	Christensenellaceae		pTF	2.21E-05	0.0127	-0.0542
176318	Christensenellaceae	Unknown	VFM	3.19E-05	0.0017	-0.0071
Collapsed	Christensenellaceae	Unknown	pTF	3.24E-05	0.0125	-0.0525
176318	Christensenellaceae	Unknown	pTF	6.52E-05	0.0140	-0.0563
4402903	Christensenellaceae	Unknown	VFM	0.0001	0.0019	-0.0074
Collapsed	Christensenellaceae		BMI	0.0008	0.0126	-0.0429
Collapsed	Christensenellaceae	Unknown	BMI	0.0012	0.0125	-0.0410
4402903	Christensenellaceae	Unknown	pTF	0.0025	0.0160	-0.0488
4402903	Christensenellaceae	Unknown	BMI	0.0081	0.0157	-0.0420
176318	Christensenellaceae	Unknown	BMI	0.0090	0.0139	-0.0369
231952	Christensenellaceae	Unknown	VFM	0.0109	0.0055	-0.0140
555547	Christensenellaceae	Unknown	VFM	0.0250	0.0040	-0.0089
555547	Christensenellaceae	Unknown	pTF	0.0356	0.0336	-0.0708
2256425	Christensenellaceae	Unknown	VFM	0.0380	0.0052	-0.0108
555547	Christensenellaceae	Unknown	BMI	0.0523	0.0329	-0.0640

Table 7: Adiposity associations with *Christensenella*

These associations are negative, indicating an increase of Christensenella in low adiposity individuals, though this does not indicate a causal relationship. Following adjustment for BMI, all associations with pTF lose nominal significance and only 4 associations with VFM remain (Table 8).

OTU	Family	Genus & Species	Pheno	P Value	Std Err	Estimate
176318	Christensenellaceae	Unknown	VFM	0.00033	0.00962	-0.0347
Collapsed	Christensenellaceae	Unknown	VFM	0.00334	0.00877	-0.0258
Collapsed	Christensenellaceae		VFM	0.00462	0.00886	-0.0252
231952	Christensenellaceae	Unknown	VFM	0.0101	0.0317	-0.0818

Table 8: Adiposity associations with Christensenellaceae following adjustment for BMI.

4.2.2 *Christensenella and inflammation*

Visceral fat tissue is able to secrete the inflammatory cytokine, Il-6 which may contribute to the widespread systemic inflammation that is associated with obesity. As the strongest *Christensenella* associations were with visceral fat, I chose to explore white blood cell (WBC)

counts in the TwinsUK cohort as WBCs are typically increased in an inflammatory state. These were measured in blood that was taken during the subject's annual clinical visit using FACS [Nalls et al., 2011].

Phenotype	Abbreviation	Description	Sample Size	Time between samples
White Blood Cell Count	WBC	A measure of the white blood cells (leukocytes) present in the blood.	913	2-6 years
Neutrophils	NA	The most abundant leukocyte in blood.		
Monocytes	NA	The largest leukocyte in blood.		
Lymphocytes	NA	Cells important for immunity.		

Table 9: White blood cell phenotypes explored.

Using a linear mixed-effects regression on approximately 900 twins, I obtained 16 nominally significant associations (P<0.05) with white blood cell count, neutrophils, lymphocytes and monocytes (Table 10).

OTU	Family	Genus & Species	Pheno	P Value	Std Err	Estimate
Collapsed	Christensenellaceae	Unknown	WBC	0.00048	0.0134	-0.0471
Collapsed	Christensenellaceae		WBC	0.00059	0.0136	-0.0469
Collapsed	Christensenellaceae		Neutrophils	0.00094	0.0138	-0.0459
Collapsed	Christensenellaceae	Unknown	Neutrophils	0.00096	0.0137	-0.0453
231952	Christensenellaceae	Unknown	WBC	0.00111	0.0494	-0.1615
Collapsed	Christensenellaceae	Unknown	Monocytes	0.00173	0.0137	-0.0431
2256425	Christensenellaceae	Unknown	Neutrophils	0.00228	0.0482	-0.1475
Collapsed	Christensenellaceae		Monocytes	0.00244	0.0139	-0.0422
2256425	Christensenellaceae	Unknown	WBC	0.00401	0.0473	-0.1365
231952	Christensenellaceae	Unknown	Lymphocytes	0.00749	0.0477	-0.1278
231952	Christensenellaceae	Unknown	Neutrophils	0.00831	0.0506	-0.1339
176318	Christensenellaceae	Unknown	WBC	0.0159	0.0151	-0.0366
231952	Christensenellaceae	Unknown	Monocytes	0.0221	0.0510	-0.1170
2256425	Christensenellaceae	Unknown	Lymphocytes	0.0264	0.0458	-0.1018
176318	Christensenellaceae	Unknown	Neutrophils	0.0311	0.0154	-0.0332
176318	Christensenellaceae	Unknown	Monocytes	0.0374	0.0154	-0.0322

Table 10: Christensenella associations with white blood cell types.

The peak association was between an unknown genus within Christensenellaceae and WBC counts (P=0.00048, β=-0.047, SE=0.0134). Once again, these associations were negative, suggesting an increase in white blood cell types when Christensenellaceae is less abundant. This would in fact support the visceral fat findings, showing an in-

crease in immune cells when *Christensenella* is reduced. Adjustment for BMI reduces the number of nominally significant associations to 12, however WBC remains the top association (P=0.004, β=-0.13, SE=0.049).

4.2.3 *Candidate gene analysis*

Christensenella is the most heritable microbe in this dataset and I was therefore interested to see if I could pinpoint any host gene variants that associate with *Christensenella* which could explain the phenotypic associations seen above. The dataset is underpowered to perform a full genome-wide association study however, so instead I conducted two candidate gene analyses. The first was to determine host adiposity gene influences on *Christensenella* and the second was to determine immune gene influences.

4.2.3.1 *Adiposity gene variant candidate gene analysis*

The NHGRI-EBI GWAS Catalog [Welter et al., 2014] is a catalog of published results from previous GWAS studies. Results can be searched according to phenotype of interest and downloaded for further use. The list of candidate obesity GWAS SNP associations was extracted from the NHGRI-EBI GWAS Catalog. The reported SNPs were taken as the lead SNP, and candidate gene regions were extended to include additional SNPs within a 10 kb region either side of the lead SNP. Overall, I considered associations with 1795 SNPs across 53 unique genomic regions. No MAF criteria were applied during the initial SNP selection. The analysis was performed using GEMMA [Zhou and Stephens, 2012] using a kinship matrix to account for family relatedness within the sample as described in Chapter 2.

OTU	Family	Genus & Species	Gene	SNP	MAF	P Value	Beta
2256425	Christensenellaceae	Unknown	BDNF	rs7934165	0.475	2.60E-05	-1.33E-01
2256425	Christensenellaceae	Unknown	BDNF	rs2030324	0.475	2.67E-05	-1.33E-01
2256425	Christensenellaceae	Unknown	BDNF	rs10767665	0.475	2.67E-05	-1.33E-01
2256425	Christensenellaceae	Unknown	BDNF	rs2049046	0.461	7.20E-05	-1.24E-01

Table 11: Adiposity candidate gene associations with *Christensenella*.

There was only one gene association with *Christensenella* that had suggestive evidence for association (threshold P= $1x10^{-4}$, Table 11). This was with variants within the *BDNF* gene (Peak P=$2.6x10^{-5}$, MAF=0.475). *BDNF* encodes the brain-derived neurotrophic factor which is found at synapses and helps to regulate synaptic plasticity. Due to the expression of *BDNF* in locations of the brain that control body weight and eating, it is believed that *BDNF* plays an important role in these activities [Lapchak and Hefti, 1992]. How this might affect *Christensenella* is unclear but one hypothesis may be that genetic variants within *BDNF* affect what the host eats, thus impacting the levels of *Christensenella*.

4.2.3.2 *Immune gene variant candidate gene analysis*

I performed a candidate gene analysis using 1039 SNPs found in 278 regions of the human genome that had previously been linked with immune function in the TwinsUK cohort [Roederer et al., 2015]. No host genetic associations with *Christensenella* surpassed a multiple testing threshold of P=$1.9x10^{-7}$. There were 159 nominally significant associations. The top association was with a variant within the *GPC6* gene (rs4773780, P=$1.09x10^{-05}$, MAF=0.139). GPC6 encodes the protein glypican-6 which has been implicated in cell growth and division [Oikari et al., 2016]. Interestingly, silencing of this gene in bile duct cell lines caused an increase of proinflammatory markers in these cells [Karlsen et al., 2010] suggesting a role for the gene in minimis-

ing inflammation. Again, how this might influence *Christensenella* is unclear.

OTU	Family	Genus & Species	Gene	SNP	MAF	P Value	Beta
231952	Christensenellaceae	Unknown	GPC6	rs4773780	0.139	1.09E-05	-1.86E-01
Collapsed	Christensenellaceae	Christensenella	ATF6	rs2341479	0.135	6.46E-04	-1.10E-01
176318	Christensenellaceae	Unknown	FAS	rs3740286	0.373	1.10E-03	-3.80E-01
2256425	Christensenellaceae	Unknown	C10orf96	rs7099441	0.106	1.41E-03	-1.66E-01
Collapsed	Christensenellaceae		MYL3	rs3792558	0.089	1.42E-03	-6.65E-01
Collapsed	Christensenellaceae	Unknown	MYL3	rs3792558	0.089	1.64E-03	-6.64E-01
Collapsed	Christensenellaceae		ENTPD1	rs12761960	0.096	2.48E-03	-6.92E-01
Collapsed	Christensenellaceae		ENTPD1	rs12770313	0.096	2.48E-03	-6.92E-01
Collapsed	Christensenellaceae		ENTPD1	rs11595122	0.096	2.48E-03	-6.92E-01
Collapsed	Christensenellaceae	Christensenella	ATF6	rs2298019	0.1	2.59E-03	-1.18E-01
Collapsed	Christensenellaceae		ENTPD1	rs3181115	0.096	2.74E-03	-6.84E-01
Collapsed	Christensenellaceae		ALDH18A1	rs11592789	0.093	2.78E-03	-6.80E-01
Collapsed	Christensenellaceae	Unknown	ENTPD1	rs12761960	0.096	2.95E-03	-6.83E-01
Collapsed	Christensenellaceae	Unknown	ENTPD1	rs12770313	0.096	2.95E-03	-6.83E-01
Collapsed	Christensenellaceae	Unknown	ENTPD1	rs11595122	0.096	2.95E-03	-6.83E-01

Table 12: Top 15 nominally significant immune gene variant associations with *Christensenella*.

The 6th ranked gene association between host immune genetic variants and *Christensenella* was observed with variants within the Ectonucleoside triphosphate diphosphohydrolase 1 (*ENTPD1*) gene that encodes the Treg Cluster of Differentiation 39 (CD39, Peak P=2.48x10^{-3}, MAF=0.096, Figure 10). Tregs, or regulatory T cells, are essential immune suppressors that are critical to distinguishing between host and non-host cells. CD39 is an important marker of inflammation. It has the ability to revert a cell back to its anti-inflamed state via degradation of ATP, ADP and AMP to adenosine [Borsellino et al., 2007]. White blood cells such as neutrophils and monocytes express CD39 acting as transports to get the TREG to key sites of inflammation and furthermore, mice deficient in CD39 are unable to suppress T-cell proliferation [Beldi et al., 2008]. The negative association here would suggest a reduction of *Christensenella* in individuals with these variants. If we assume that these variants impact the functionality of CD39 resulting in increased inflammation, then these results would also concur with the findings presented in Table 10 and support the link between *Christensenella* and inflammation.

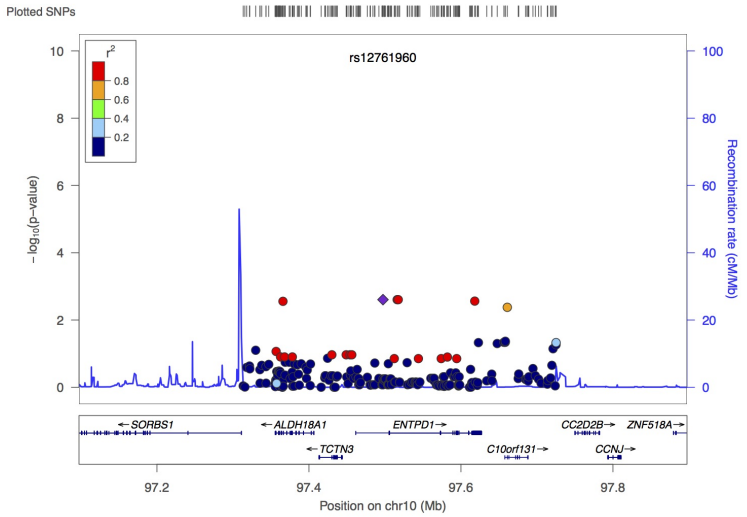

Figure 10: Locus-zoom plot of ENTPD1 associations with *Christensenella*

4.3 CONCLUSION

My aim in this chapter was to determine whether *Christensenella* was associated with adiposity or body mass while also developing a hypothesis about how *Christensenella* affects adiposity. The abundance of *Christensenella* associates with visceral fat, the component of adipose tissue with important metabolic implications, independently of BMI. Candidate gene analysis exploring obesity genes found a nominally significant association between *Christensenella* and BDNF, a gene responsible for signalling and synaptic plasticity in the brain, particularly in regions that control food and drink consumption [Bramham and Messaoudi, 2005]. An immediate theory as to how this gene influences *Christensenella* is not apparent, but one hypothesis is that variations in BDNF may influence what the host eats or how much [Rios, 2013], thus impacting levels of *Christensenella*.

As visceral fat has important implications with inflammation and immune response, I pursued association analyses between *Christensenella* and immune phenotypes available in the TwinsUK cohort at the time. *Christensenella* was significantly associated with a

number of leukocyte sub-types as well as overall white blood cell count. White blood cells are good markers of inflammation due to their involvement in both the innate and adaptive immune response, and transport of Tregs. Furthermore, another candidate gene analysis of immune traits showed *Christensenella* to be associated with host genetic variants in *GPC1* and *ENTPD1*, both of which have implications with inflammation. While it is unclear how *GPC1* may interact with *Christensenella*, *ENTPD1* encodes an anti-inflammatory Treg that is typically present on neutrophils and monocytes. It is potentially via leukocytes that CD39 interacts with *Christensenella* and thus results in the associations seen not only in the immune candidate gene analysis, but also the white blood cell association analysis.

This suggests a potential interaction between *ENTPD1 Christensenella* and inflammation that warrants further investigation.

The approaches used in this section yielded interesting results at the candidate level. In Chapter 5 I apply these approaches on a microbiome-wide scale in order to characterise the gut microbiome in obesity.

5

MICROBIOME-WIDE ASSOCIATION STUDY (MIWAS) OF OBESITY

5.1 BACKGROUND

Of particular interest to researchers is the interaction of the microbiota with disease. Many diseases have now been linked to changes in the gut microbiota and these include ulcerative colitis [Rooks et al., 2014], obesity [Turnbaugh and Gordon, 2009, Walters et al., 2014], diabetes [Kostic et al., 2015] and irritable bowel syndrome [Major and Spiller, 2014]. A number of study designs have been used to understand the role of the microbiome in these diseases such as mouse models, twin designs, case-controls and statistical modelling and in this chapter I will focus on the methods that I have used, twin design and statistical modelling.

Due to the large number of phenotypes characterised in the TwinsUK cohort, I performed a number of association scans comparing each of the phenotypes with the entire 16s gut microbiota. This I have termed a Microbiome-Wide Association Study (MiWAS).

5.1.1 *Obesity*

Obesity has become a significant problem in recent years, originally within Western populations but now extending worldwide. In 2008 it was estimated that approximately between 10-14% of the world's population was obese, posing an enormous burden on healthcare costs

(http://www.who.int/gho/ncd/risk_factors/obesity_text/en/). Despite this, obesity is largely preventable with the diet and lack of exercise being the largest risk factors for the development of the disease [Barness et al., 2007, Kelly et al., 2008]. There has been a large drive to understand the role of the microbiome in obesity and there are many differing views on whether differences observed are cause or effect.

Studies on early life exposure to antibiotics have found some evidence suggesting disruption of the microbiota may result in obesity later in life [L et al., 2014, Trasande et al., 2013]. Mouse models have also provided some evidence to suggest a causative role for the microbiome in obesity. Germ-free mice innoculated with the gut microbiome from conventionally-raised mice show a 60% increase in adiposity [Backhed et al., 2004] while faecal analysis of obese mice shows differences in community profiles from lean mice [Ley et al., 2005]. But of course, there are studies that show no differences in the gut microbiome between lean and obese subjects [Duncan et al., 2008]. Studies do seem to agree on the same phyla of bacteria that change in obesity, observing increases in Firmicutes and decreases in Bacteroidetes in lean/obese individuals, in what has come to be known as the Firmicutes:Bacteroides ratio.

A recent meta-analysis by the Knight Lab of available microbiome studies of obesity that have focused on BMI showed no consistent trend across microbiome studies, in either microbial profiles or alpha diversity [Walters et al., 2014]. The authors propose that this is likely due to the differences in sample collection and methodological differences between studies, such as the use of qPCR vs amplicon sequencing. They highlight the importance of developing a "gold-standard" pipeline for handling of microbiome samples, especially if we expect to replicate findings as is now common practice in other fields.

Measurement of obesity in microbiome studies has been mostly limited to BMI in humans. As BMI measures overall adiposity without distinction between fat mass and lean mass it is not the most accurate measure of obesity [Romero-Corral et al., 2008]. Mouse studies however tend to use epididymal fat weight, a visceral fat proxy in mice, and DEXA-derived measurements of total fat mass, both of which are a far more accurate assessment of adiposity. DEXA (dual-energy x-ray absorptiometry) uses x-rays to assess body composition. This procedure works on the basis that x-rays travel through different tissues at differing rates. Measuring the transit time of these rays can provide a whole-body view of the location of fat and in what quantity. Differences in this characterisation of obesity could in part explain the differing results found by the multitude of obesity studies. Another point worth noting is the lack of studies using visceral fat as a marker of obesity. Visceral fat, or adipose tissue, is the fat tissue that is highly associated with cardio-metabolic disorders [Sironi et al., 2012]. It is the deep layer of fat tissue that sits around organs and is the hardest to be rid of once it has developed. Traditionally however, visceral fat has been difficult to measure in humans, requiring time-consuming and expensive CT and MRI scans. Recent developments in DEXA technology means that visceral fat can now be estimated in the same location as a typical CT slice cross-section in the lumbar region L4-L5. DEXA scans are much quicker than CT scans and also estimate overall body fat as well as android, gynoid and trunk fat. A number of studies have confirmed that DEXA estimations of visceral fat are highly correlated with CT scan measurements [Snijder et al., 2002], making DEXA measures a viable alternative in studying obesity.

In this Chapter my aim was to investigate the association between the gut microbiome and obesity using multiple measures of adiposity in a large sample twins.

5.2 RESULTS

In order to investigate the effect of the gut microbiome on obesity in my dataset, I used DEXA-derived measurements on subjects who had provided a faecal sample for 16s profiling. Summary characteristics of the obesity phenotypes can be found in Table 13. The sample was composed of a total of 960 predominantly female twins as described in Chapter 2.

Phenotype	Average (Range)	Sample Size	Abbreviation
Visceral Fat	567.21 (1.28 - 1919)	866	VFM
% Trunk Fat	31 (3.2 - 54.7)	960	pTF
BMI	29 (15.7 - 50.7)		BMI
Android Fat	2152.4 (295.67 - 6590.99)	866	AFM
Android Lean	2989.48 (1671.46 - 5863.63)		ALM
Android Total Mass	40.13 (58.6 - 11.84)		ATM
Gynoid Fat	4596.55 (1033 - 10556.22)		GFM
Gynoid Lean	6167.25 (3600.15 - 10335.66)		GLM
Gynoid Total Mass	42. 14 (59.81 - 15.57)		GTM
Android:Gynoid Ratio	0.46 (0.16 - 1.08)		A:G
Subcutaneous Fat	1611.63 (416.39 - 3621.28)		SFM

Table 13: DEXA variables summary

5.2.1 *Heritability of Adiposity*

First, I estimated heritability of all adiposity measures used, including the new visceral measures using OpenMX. Previous heritability estimates of other adiposity measures, including BMI, had been quite

high, attributing approximately 70% of the variance in obesity to genetic effect [Chaput et al., 2014, Fox et al., 2007]. In this dataset, I estimated that BMI has a heritability of 76% and % Trunk Fat has a heritability of 57%, so it is unsurprising that visceral fat has a heritability of 73% (Figure 11). This value is greater than estimates in previous studies of visceral fat heritability that show heritability to lie between 36-55% [Fox et al., 2007, Chaput et al., 2014, Rice et al., 1997] .

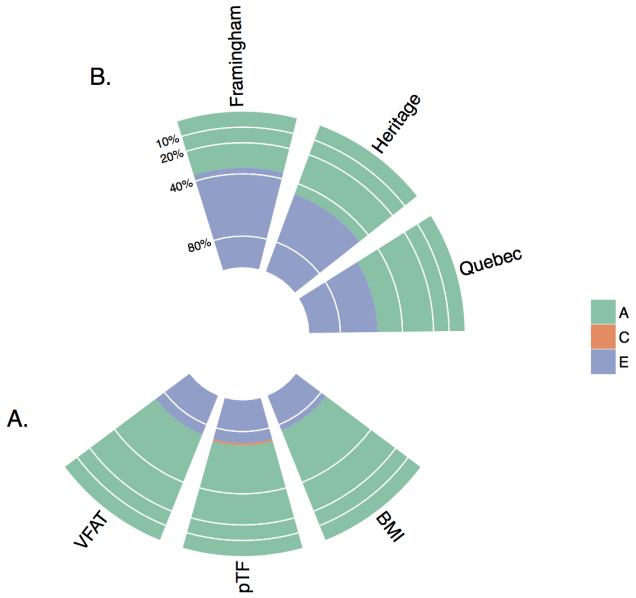

Figure 11: Polar histogram of adiposity heritability in the TwinsUK cohort. Green indicates additive genetics, orange indicates common environment and blue indicates unique environment. A. Heritability estimates of visceral fat (VFAT), % Trunk Fat (pTF) and BMI in the TwinsUK cohort. B. Estimates of visceral fat in the Framingham [Fox et al., 2007], Heritage [Rice et al., 1997] and Quebec [Chaput et al., 2014] cohorts.

5.2.2 *Microbial Diversity*

The next analyses were concerned with determining microbial diversity changes with changes in adiposity, to determine if microbial diversity is reduced in those subjects with more adipose tissue. Alpha

diversity, as measured by Faith's diversity or PD whole tree averages, was calculated using QIIME by Ms Goodrich. PD whole tree averages calculates diversity based upon branch lengths in the phylogenetic trees. OTUs that are closely related to other OTUs add a small amount to the diversity while OTUs that are entirely unique add a large amount to the diversity. This differs to other methods of alpha diversity, such as Shannon diversity, as it does not account for species abundance.

Alpha diversity showed a significant negative association with each of the fat phenotypes. However no significant association was obtained with either of the lean phenotypes, ALM and GLM (Figure 12). Visceral fat mass showed the most significant association ($P=5.11x10^{-6}$). This confirms the findings in previous studies and the validity of the current data.

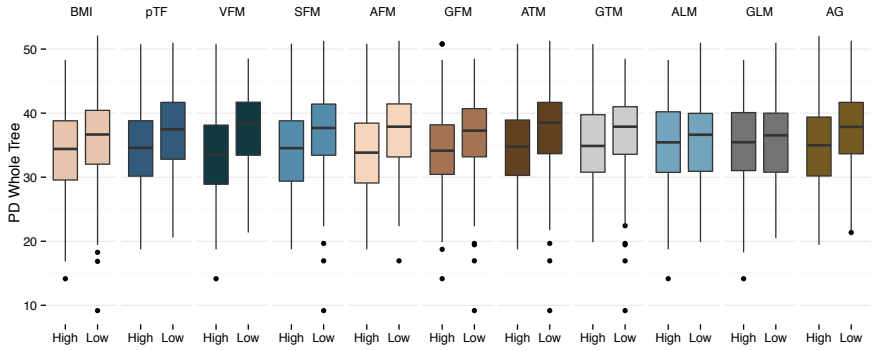

Figure 12: Boxplots showing reduced alpha diversity between high and low fat subjects. BMI high and low cut-offs were determined using the clinical standard healthy BMI ranges (>30 = High, <25 = Low). For all other measures High and Low were determined as being >1SD from the mean.

Two recent metagenomics studies have both found lower microbial gene counts in obese individuals indicating a reduction in microbial diversity [Le Chatelier et al., 2013, Cotillard et al., 2013], supporting the findings above. In Cotillard et al. [2013], a dietary intervention in obese individuals was able to increase the microbial gene

count in obese individuals providing a potential treatment option for microbiota restoration in obesity.

Beta diversity showed no clustering according to status of obesity (Figure 13).

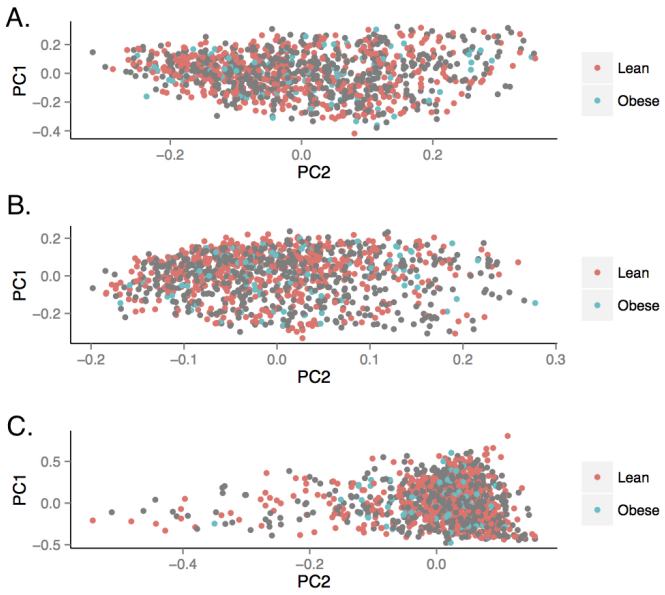

Figure 13: Scatterplots coloured according to obesity status as measured by BMI. A. Bray-Curtis metric. B. Unweighted Unifrac metric. C. Weighted Unifrac metric.

5.2.3 *Microbial Associations With Adiposity*

I next sought to determine which individual microbes differed significantly in abundance in obesity. Using a linear mixed effects model, defining the phenotype as the response variable and the OTU as a predictor, adjusting for zygosity, sex, batch and family structure (refer to Chapter 2 for more detail), I performed 9999 tests, comparing 11 phenotypes against 909 microbial units (768 OTUs and 141 taxonomy levels). There were 66 associations that surpassed a Bonferonni threshold of $P=5x10^{-6}$ and altogether, 2981 that surpassed nominal significance (Figure 14).

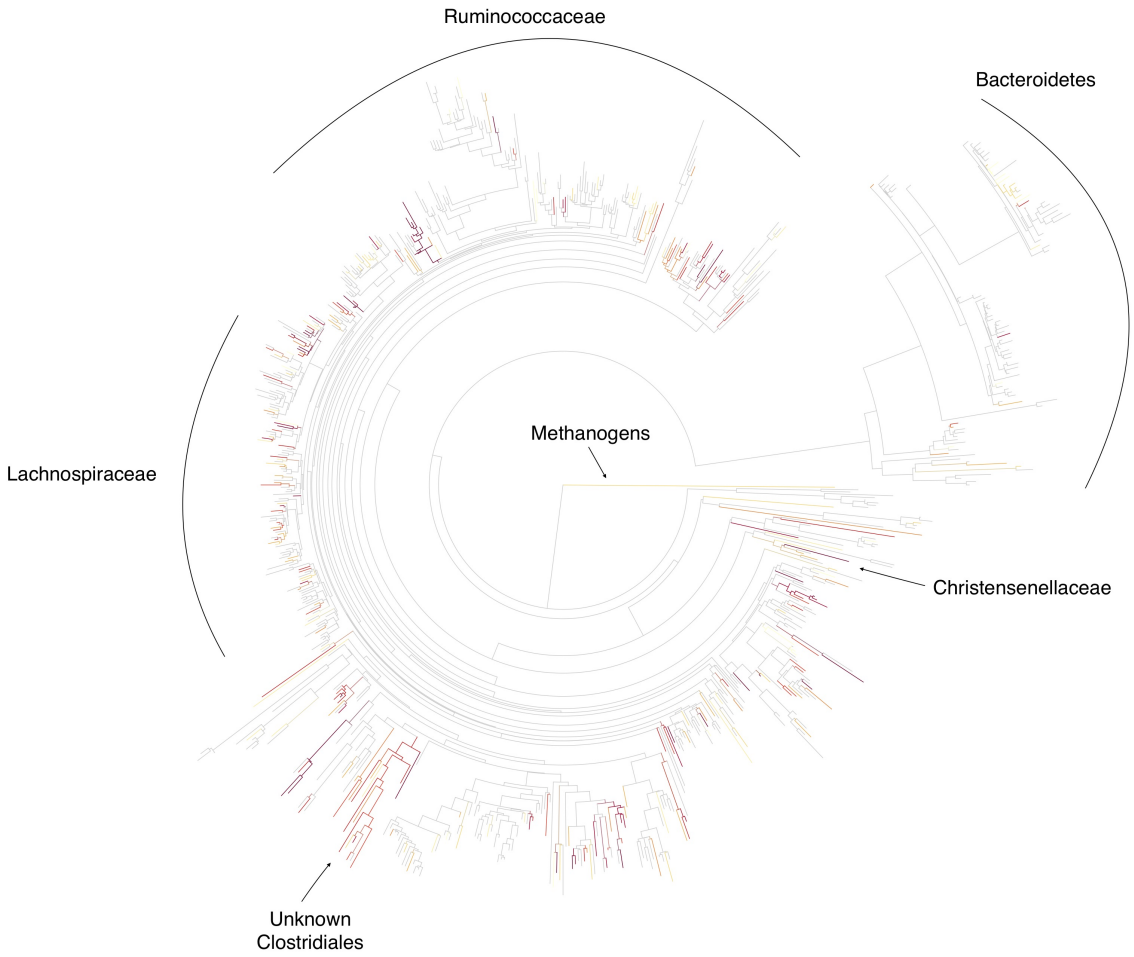

Figure 14: Phylogenetic tree of OTUs associated with visceral fat in the TwinsUK cohort, coloured according to significance of change. Yellow indicates nominal significance and red the most significant.

If we consider a conservative Bonferroni threshold of $P=5x10^{-6}$, 66 associations pass multiple testing (Table 14 and these were associations between 34 microbial units (20 OTUs and 14 taxonomy levels) and the 11 adiposity phenotypes. A complete table of the 66 associations can be found in Appendix A (Table A2). 38% of these associations are with visceral fat showing the strength of this variable in finding microbial associations with obesity.

OTU	Parent Taxonomy (Level)	Taxonomy (Level)	Pheno	P Value	SE	Beta
207487	Lachnospiraceae (Family)	Lachnospira (Unknown Species)	VFM	1.82E-09	0.0053	-0.0324
199344	Lachnospiraceae (Family)	Unknown (Genus & Species)	GTM	3.46E-08	0.0414	0.2352
199344	Lachnospiraceae (Family)	Unknown (Genus & Species)	GLM	4.46E-08	0.0173	0.0971
4349261	Lachnospiraceae (Family)	Unknown (Genus & Species)	pTF	9.09E-08	0.0342	-0.1844
207487	Lachnospiraceae (Family)	Lachnospira (Unknown Species)	AFM	2.08E-07	0.0173	-0.0907
Collapsed	Tenericutes (Phylum)		pTF	2.24E-07	0.0088	-0.0463
289734	Lachnospiraceae (Family)	Unknown (Genus & Species)	VFM	2.30E-07	0.0014	0.0071
176269	Lachnospiraceae (Family)	Lachnospira (Unknown Species)	VFM	2.34E-07	0.0014	-0.0075
199344	Lachnospiraceae (Family)	Unknown (Genus & Species)	BMI	2.42E-07	0.0560	0.2957
193769	Lachnospiraceae (Family)	Ruminococcus gnavus	ATM	2.45E-07	0.0167	0.0879
193769	Lachnospiraceae (Family)	Ruminococcus gnavus	GLM	2.85E-07	0.0095	0.0497
193769	Lachnospiraceae (Family)	Ruminococcus gnavus	GTM	2.88E-07	0.0226	0.1197
Collapsed	Tenericutes (Phylum)	Mollicutes (Class)	pTF	5.17E-07	0.0098	-0.0501
322835	Lachnospiraceae (Family)	Lachnospira (Unknown Species)	VFM	6.00E-07	0.0054	-0.0272
199344	Lachnospiraceae (Family)	Unknown (Genus & Species)	ATM	6.19E-07	0.0308	0.1577
Collapsed	Tenericutes (Phylum)		VFM	6.76E-07	0.0011	-0.0053
Collapsed	Mollicutes (Class)	RF39	pTF	7.77E-07	0.0105	-0.0527
Collapsed	RF39 (Order)	Unknown (Family)	pTF	7.77E-07	0.0105	-0.0527
Collapsed	RF39 (Order)	Unknown (Family & Genus)	pTF	7.77E-07	0.0105	-0.0527
Collapsed	RF3 (Class)		VFM	8.13E-07	0.0035	-0.0177

Table 14: Top 20 MiWAS results with adiposity measures.

The main families with taxa that had high significance were Lachnospiraceae and Ruminococcaceae. The top association was between a *Lachnospira* OTU and visceral fat mass (ANOVA, $p=1.82x10^{-9}$, $\beta= -0.03$). There were a number of *Lachnospira* OTUs in the results that surpassed Bonferroni correction as well as a number of Tenericute collapsed taxonomies. The families Ruminococcaceae and Lachnospiraceae are members of the Firmicutes and have previously been associated with obesity [Daniel et al., 2014, Zhang et al., 2009]. A significant number of OTUs that contributed to the top significant associations were moderately heritable including the family, Christensenellaceae, known previously to have associations with BMI [Goodrich et al., 2014] as well as significant associations with visceral fat mass (Chapter 4, Table A1)($p=3.51x10^{-6}$).

To date, this is the first visceral fat-microbiome study in humans and the results have implications for the role of the microbiota in obesity. Visceral fat has important ramifications on cardiovascular disease and studies have shown it to be implicated in insulin resistance by secreting retinol binding protein 4 (RBP4) [Klöt-

ing et al., 2007]. Furthermore, visceral fat promotes the inflammatory phenotype often seen in obesity by secreting the proinflammatory adipokine, IL-6 [Fontana et al., 2007] which can further lead to insulin resistance and type 2 diabetes. Due to the highly heritable status of visceral fat and aspects of the human microbiota, in the next section I will explore potential host adiposity gene effects on the gut microbiome association in obesity.

5.2.4 *Human Obesity-Linked Genes and the Gut Microbiome in Obesity*

Obesity is highly heritable, as shown in Section 5.2.1, and some aspects of the microbiome have previously been shown to be heritable (see Chapter 3). With this in mind, I performed a candidate gene analysis (CGA) taking genetic variants that had previously been associated with obesity and testing their association with each OTU and collapsed taxonomy that had significantly associated with obesity within the TwinsUK 16S dataset.

At the time of thesis submission, there are 97 loci associated with obesity from genome-wide association studies [Locke et al., 2015]. At the time of analysis however, there were only 32 "gold-standard" loci that had been found to be associated with obesity [Speliotes et al., 2010]. There were an additional 21 loci reportedly associated with BMI in the NHGRI-EBI GWAS Catalog at the time. The analyses presented in this chapter are therefore based on this total of 53 obesity-loci. Shortly before thesis submission, the analyses in this section were repeated and the results with the latest list of 97 obesity loci and expanded microbiome dataset are shown in the appendix (Appendix A, Table A3 and Appendix B, Section B.1).

Of all loci associated with obesity in humans to date, two human obesity genes are of particular interest, *FTO* and *MC4R*. *FTO* was

the first obesity locus discovered in 2007 [Frayling et al., 2007] and is widely regarded as one of the biggest contributors to the obesity epidemic. It is now believed that *FTO* affects obesity by controlling energy intake as opposed to expenditure [Speakman et al., 2008] and in 2013 this was further supported when it was found that genetic variants in *FTO* affect levels of the hormone ghrelin which controls satiation and food intake [Karra et al., 2013]. But a recent study has proposed that *FTO* actually only has a very small, peripheral effect on obesity and that it is in fact *IRX3*, a neighbouring gene that is the culprit [Smemo et al., 2014]. The authors suggest that *FTO* interacts with the promoters of the *IRX3* gene resulting in the obese phenotype.

Melanocortin 4 receptor (*MC4R*) encodes a protein responsible for binding the hormone, α-melanocyte stimulating hormone. It has been found to possess very strong effects on obesity. *Mc4r*-deficient mice were found to become extremely obese [Huszar et al., 1997] and the prevalence of MC4R mutations in individuals with a BMI >30 is approximately 2.5% [Larsen et al., 2005]. In addition, a number of genome-wide association studies have determined associations between BMI and genetic variants within or near*MC4R* [Scherag et al., 2010, Loos et al., 2008, Chambers et al., 2008].

In this section I tested for genetic association between human genetic variants at 53 20kb regions that had been reported to be associated with BMI in the NHGRI-EBI GWAS Catalog, selecting the lead obesity-associated GWAS SNP as the centre of each candidate gene region. Altogether, there were 1795 human SNPs in these 53 regions and a test of association was performed for each of these variants and all 34 obesity-associated OTUs and taxonomy levels identified in Section 5.2.3 above.

Three loci that had previously been associated with obesity showed nominally significant associations with at least one of the

34 adiposity-associated microbes identified in Section 5.2.3 (listed in Table 14 and Table A2). Only the top 3 associations passed a multiple-testing threshold of P=2.7x10^{-5} (0.05/(34*53)). The peak association was between a Lachnospiraceae OTU (OTU 174911) and a variant within the *RPTOR* gene (rs8081087, P=1.44x10-7, Figure 15).

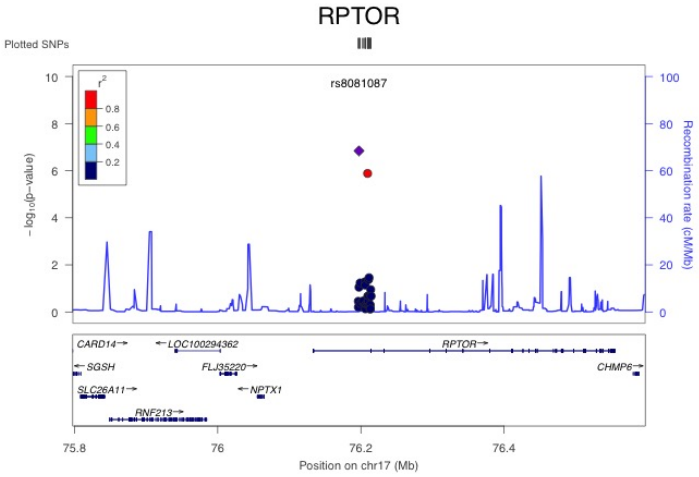

Figure 15: Locus-zoom plot of the association between rs8081087 and Lachnospiraceae

RPTOR is responsible for controlling the insulin-signalling pathway through the mTOR pathway in response to nutrient availability. When there is an abundance of nutrients, RPTOR binds and activates mTOR which in turn promotes the kinase activity S6K1, leading to phosphorylation of IRS-1 and subsequent inhibition of insulin signalling. Two other gene variants surpassed nominal significance but not the multiple-testing threshold. These were genetic variants in *PROX1* and *TFAP2B*.

I then explored the association between rs8081087 with visceral fat, pTF, BMI and insulin, and also explored adiposity associations with the significant Lachnospiraceae OTU 179411. While the SNP rs8081087 was not significantly associated with adiposity phenotypes, other variants within *RPTOR* showed nominally significant associations with VFM (peak P=8.37x10^{-4}) as well as insulin (peak

P=9.93x10^{-4}). Furthermore, the Lachnospiraceae OTU, 174911, was nominally associated with the serum metabolite lathosterol (P=0.000185), a precursor to cholesterol whose levels have been shown to differ during insulin resistance [Pihlajamaki et al., 2004]. Further metabolite association results will be discussed in detail in Chapter 6. In addition, gene expression profiles across multiple tissues in the pilot GTEX data showed that RPTOR is expressed at greater levels in sigmoid colon, oesophagus (gastroesophageal junction and muscularis), and skeletal muscle (http://www.gtexportal.org/).

5.2.5 *Exploring Causality*

In order to infer causality, I decided to apply Mendelian Randomisation (MR) to the data, specifically exploring the RPTOR SNP - Lachnospiraceae OTU 174911 - Visceral fat pair-wise association results. To this end I performed analyses in collaboration with Dr Emily Davenport and Dr Andrew Clark at Cornell University, and the final MR results presented in Figure 16 were calculated by Dr Davenport. Mendelian randomisation is a technique to test for causal relationships using SNP data as an anchor to test whether an intermediate phenotype causes the phenotype of interest [Lawlor et al., 2008]. There are 3 major assumptions of mendelian randomisation (Figure 16). The first is that the genotype of interest is associated with the intermediate phenotype. The second assumption is that the genotype of interest has no relationship to the confounding factors that may affect the relationship between the intermediate phenotype and phenotype of interest. The final assumption is that the genotype is only linked to the phenotype of interest via its association with the intermediate phenotype (Figure 16).

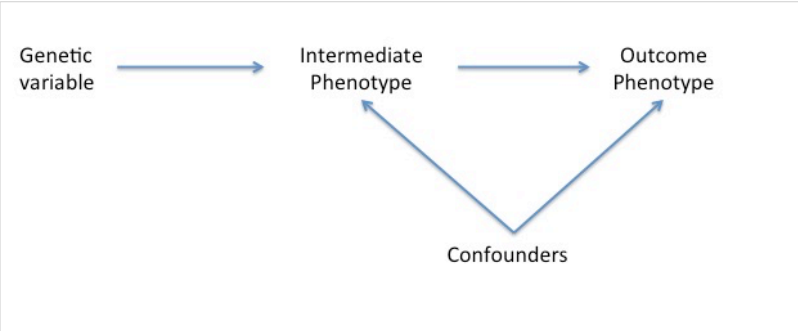

Figure 16: Diagram to show the relationship between variables assumed for mendelian randomisation.

We hypothesised that the SNP rs8081087 may influence the Lachnospiraceae OTU of interest, OTU 174911, which may in turn affect obesity. MR was done using a two-stage least squares regression, with either pTF, insulin, VFM, BMI or lathosterol as the outcome variable. The beta coefficients for each phenotype were plotted in a forest plot (Figure 17) and with the exception of insulin, the phenotypes had very small effect sizes. If OTU 174911 was mediating the effect of the RPTOR SNP on visceral fat I would expect to see a significant shift in the regression coefficient away from zero. Ultimately, the analyses were not significant and causality could not be inferred for any of the phenotypes.

The lack of significance however, may simply be down to a lack of power. Power to detect associations in genetic studies is often a concern as any given genetic variant often explains only a small proportion of the phenotypic variance. An online tool, mRnd, created by the Visscher lab, calculates the sample size one would require to assess causality using a 2-stage least squares randomisation at a given power threshold [Brion et al., 2013]. Using this tool, and assuming the association between rs8081087 and OTU 174911 explained 1% of the variance, I was able to calculate the sample size required to reach 80% power to detect causality (Figure 18).

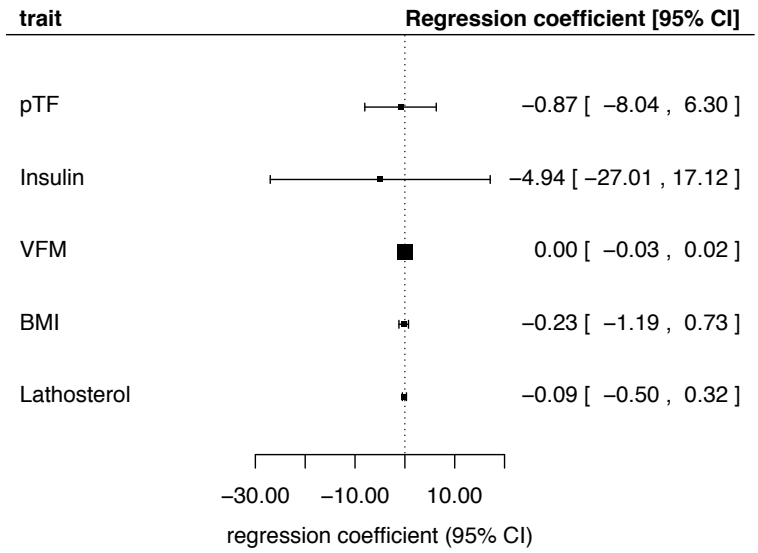

Figure 17: For each obesity trait, the forest plot shows the betas and confidence intervals from Mendelian Randomisation.

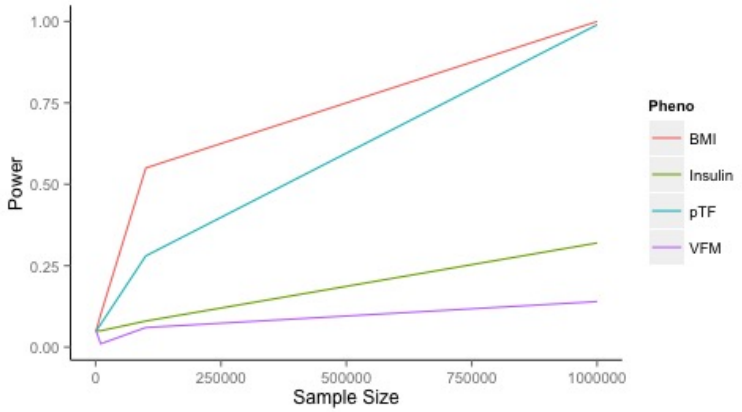

Figure 18: Power curves for determining sample size for Mendelian Randomisation on obesity traits and OTUs.

The results show that using mendelian randomisation, we would require hundreds of thousands of subjects to detect causality and it is therefore unsurprising that we do not see an association with our sample of 960 individuals. We can assume that this is in fact a generous estimate as our 1% assumption is based upon the FTO SNP which explains approximately 1% of the variance observed in BMI [Frayling et al., 2007].

5.3 CONCLUSION

In this chapter I have outlined microbiome associations with the novel phenotype, visceral fat and 9 other adiposity phenotypes. I observed that the strongest associations between the gut microbiome in obesity were obtained for visceral fat, which is the adiposity phenotype that is most relevant for cardiometabolic disease risk. I have also confirmed the heritability estimates for multiple adiposity phenotypes within this cohort, showing visceral fat to possess a high heritability of 73%. I have also confirmed the association between microbial diversity and the gut microbiome in this dataset. Using host genotype, I have found a promising gene candidate that may influence the human microbiome and while not successful, attempted to use mendelian randomisation to infer a causal relationship between this gene and obesity, with the microbiome at a mediator.

The human gut microbiome plays a large role in obesity, but it is unclear so far whether this role is causal or reactive. While mouse models may hint at a causative role, this has yet to be proven in humans. Host genetics appear to play a role in the presence of certain microbes, but the genes behind this influence are yet to be discovered. Until large study numbers can be acquired, a candidate gene approach may be the best method to seek host genes that may be interacting with host microbes.

After these analyses presented in this chapter had been completed I obtained additional 16S data just prior to thesis submission. I repeated the analyses presented in this chapter in the larger twin dataset and I summarize the updated results briefly in the Appendix B.

6

METABOLOMICS AND THE MICROBIOME

6.1 BACKGROUND

Metabolomic profiling is a useful technique to give a snapshot of metabolic functions in a tissue. While techniques such as proteomics and gene expression may reveal which gene products are being produced in any given cell, metabolomic profiling provides an overview of the functional processes within that cell. This versatile profiling technique can be performed on a number of sample types including saliva, faeces, urine and blood. Typically, the metabolites are isolated from the sample and then analysed using one of a number of methods, including nuclear magnetic resonance (NMR) and mass spectrometry (MS), to determine the identity of each metabolite.

Changes in the metabolic profile can be linked with disease. In the TwinsUK cohort in particular, a number of metabolomic findings have previously been published based on blood metabolite measurements in approximately 6000 twins. In 2013, C-glycosyl tryptophan (C-glyTryp) was found to strongly correlate with ageing and birth weight in a subset of over 6000 twins ($P = 7x10^{-157}$ and $P = 1.8 \times 10^{-8}$, respectively) [Menni et al., 2013b]. In the same cohort, 3-methyl-2-oxovalerate was found to be the strongest biomarker of impaired fasting glucose ($P = 8.46x10^{-9}$) [Menni et al., 2013a]. A genome-wide association study has also been performed on the metabolomics data in this cohort, resulting in a comprehensive database of hundreds of genetic associations [Shin et al., 2014].

Metabolites are produced not only by human cells, but also microbial cells. Penicillin for example, perhaps one of the most useful drugs of the 21st Century, is a secondary metabolite of the fungi genus *Penicillium*. In fact, a number of microbial secondary metabolites are antibiotics, necessary to compete for resources and space. Metabolon Inc have estimated which host metabolites may be produced via microbial metabolism (Figure 19).

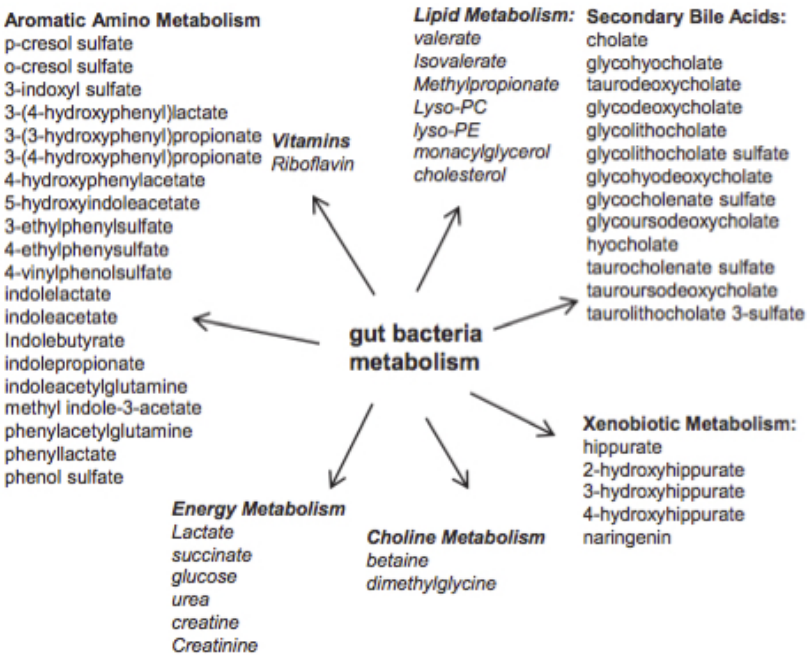

Figure 19: Host metabolites that are modulated by gut microbial metabolism. Italicised metabolites are affected by both host and microbial cells. Non-italicised metabolites are modulated by microbial cells the majority, if not exclusively, of the time. Graphic adapted from Guo et al (2015)[Guo et al., 2015].

In the absence of metagenomic and transcriptomic data to give an overview of collective microbial function, metabolomics can provide an alternative and useful insight into microbiota metabolic function. The Human Metabolome Database (HMDB, http://www.hmdb.ca/) is a comprehensive repository of metabolites found in a range of human tissues, allowing the researcher to refine a search based upon the origin of the metabolite. In this way, a search can be refined to just microbial metabolites such as phylloquinone (vitamin K1). However,

the line between host metabolite and microbial metabolite becomes blurred when we consider that microbes can produce many of the same metabolites as the host (Figure 19).

In this Chapter I sought to determine gut microbial associations with serum, plasma and faecal metabolites in a set of healthy twins. Then using a candidate approach, I was able to identify microbes that may potentially influence chronic kidney disease through production of metabolites implicated in the etiology of the disease.

6.2 RESULTS

6.2.1 *Microbiome-wide Metabolome-wide Association Study*

The sample size used in this section is 831 individuals that had both serum/plasma metabolite samples and 16s gut microbiota samples. In total, 909 microbial units and 508 metabolites (of which 311 metabolite possess a known identity) were considered. Please refer to Chapter 2 for a complete description of the data analysed within this chapter. To determine potential microbially-driven metabolites it was necessary to produce a reference association table, outlining associations between OTUs and human serum and plasma metabolites. Using a linear mixed-effects model (as outlined in Chapter 2), in this dataset of 831 individuals, metabolites were considered as the dependent variable and the OTU or collapsed taxonomy as the predictor. Covariates included BMI, batch, twin structure and zygosity.

Altogether, 111 associations passed Bonferroni correction ($P = 1.07x10^{-7}$, Table A4) and 361 metabolite-microbiota associations surpassed FDR 1% threshold, composed of 135 unique OTUs and taxonomies and 23 unique metabolites. Palmitate was the most abundantly associated metabolite, followed by cholesterol and phenylala-

nine. Saccharin, 4-acetamidobutanoate, histidine and isoleucine each had 2 associations, while hypoxanthine had 1 association (Table 15). The heritable family Christensenellaceae (collapsed family) featured in the top hits, first with palmitate ($P=3.44x10^{-26}$) and second with phenylalanine ($P=3.6x10^{-17}$). The positive beta of both these interactions suggests an increase of Christensenellaceae when palmitate increases. However this does not necessarily indicate a causal interaction.

Metabolite	N° of OTU Assocs
palmitate (16:0)	63
phenylalanine	11
cholesterol	21
saccharin	2
phenylacetate	7
hypoxanthine	1
histidine	2
isoleucine	2
4-acetamidobutanoate	2

Table 15: Bonferonni-significant metabolites and the number of Bonferonni-significant associations they had with adiposity.

6.2.2 *Metabolite-Microbe Associations In Obesity*

Of the 111 metabolite associations that passed Bonferroni correction, 16 were with OTUs and collapsed taxonomies that had previously been associated with adiposity phenotypes that surpassed Bonferonni correction (Chapter 5, Table 16). These associations were composed of 11 unique OTUs/taxonomies and 5 metabolites.

These metabolite-microbe associations were with palmitate, phenylalanine, histidine, isoleucine and 4-acetomidobutanoate. Palmitate, phenylalanine, histidine and isoleucine all have previously been

OTU	Parent Taxonomy (Level)	Taxonomy (Level)	Metabolite	P Value	Estimate	SE
Collapsed	Christensenellaceae (Family)	Unknown Genus	palmitate	2.14E-25	0.128803141	0.011953116
Collapsed	Christensenellaceae (Family)	Unknown Genus	phenylalanine	9.42E-17	0.114106241	0.013426764
Collapsed	Dehalobacteriaceae (Family)		palmitate	1.16E-14	0.211281463	0.026854464
Collapsed	Dehalobacteriaceae (Family)	Dehalobacterium (Genus)	palmitate	3.86E-14	0.21327813	0.027666538
289734	Lachnospiraceae (Family)	Unknown Genus and Species	palmitate	5.01E-14	-0.087332116	0.011361762
3195723	Ruminococaccaceae (Family)	Oscillospira (Unknown Species)	palmitate	9.31E-14	0.292033643	0.038521203
292735	Lachnospiraceae (Family)	Blautia (Unknown Species)	palmitate	2.07E-13	-0.18368881	0.024498402
3014082	Lachnospiraceae (Family)	Blautia (Unknown Species)	palmitate	1.29E-10	-0.162315269	0.024888197
4465907	Lachnospiraceae (Family)	Blautia (Unknown Species)	palmitate	4.70E-09	-0.284319785	0.047592053
289734	Lachnospiraceae (Family)	Unknown Genus and Species	phenylalanine	5.57E-09	-0.074781969	0.012627503
Collapsed	Christensenellaceae (Family)	Unknown Genus	histidine	9.26E-09	0.076426197	0.013171684
292735	Lachnospiraceae (Family)	Blautia (Unknown Species)	phenylalanine	1.18E-08	-0.157675999	0.027208389
Collapsed	Christensenellaceae (Family)	Unknown Genus	isoleucine	8.82E-08	0.071854811	0.013312587
194488	Ruminococaccaceae (Family)	Unknown Genus and Species	palmitate	9.22E-08	0.276643443	0.051318572
Collapsed	Clostridiales (Order)	Unknown Family	4-acetamidobutanoate	1.05E-07	0.045231816	0.00840111
Collapsed	Clostridiales (Order)	Unknown Family and Genus	4-acetamidobutanoate	1.05E-07	0.045231816	0.00840111

Table 16: Overlap between microbes associated with metabolites (Bonferonni) and obesity (Bonferonni).

linked to metabolic disorders. 4-acetamidobutanoate was one of a group of metabolites able to distinguish between under-nourished mice and fully-nourished mice in urine [Preidis et al., 2014]. Given the metabolic significance of these metabolites in obesity, it is interesting to see obesity-associated microbes showing associations here also. Of course, these metabolic changes may be a result of dietary changes in the host. Palmitate is the most abundant of all the fatty acids, and this may be highly associated with the microbiota simply due to this. The metabolite data is fasting data however, potentially minimising any effect from immediate dietary intakes.

6.2.3 *Metabolite-Set Enrichment Analysis*

Over representation analysis (ORA) is a useful method to determine pathway enrichment in a given set of significant metabolites. This analysis was performed using the online tool MSEA (http://www.msea.ca/) [Xia and Wishart, 2010] which uses a hypergeometric test to determine whether the set of metabolites is represented more than by chance. In order to determine if there were any pathways enriched in these obese-microbe metabolites, I took the metabolites that com-

prised the nominally significant associations from the obese microbe crossover results in Section 6.2.2 (Table 17).

Query	Match	HMDB
palmitate	Palmitic acid	HMDB00220
phenylalanine	L-Phenylalanine	HMDB00159
histidine	L-Histidine	HMDB00177
isoleucine	L-Isoleucine	HMDB00172
4-acetamidobutanoic acid	4-Acetamidobutanoic acid	HMDB03681
phenylacetylglutamine	Alpha-N-Phenylacetyl-L-glutamine	HMDB06344
indolelactate	Indolelactic acid	HMDB00671
4-methyl-2-oxopentanoate	Ketoleucine	HMDB00695

Table 17: Metabolites used for MSEA.

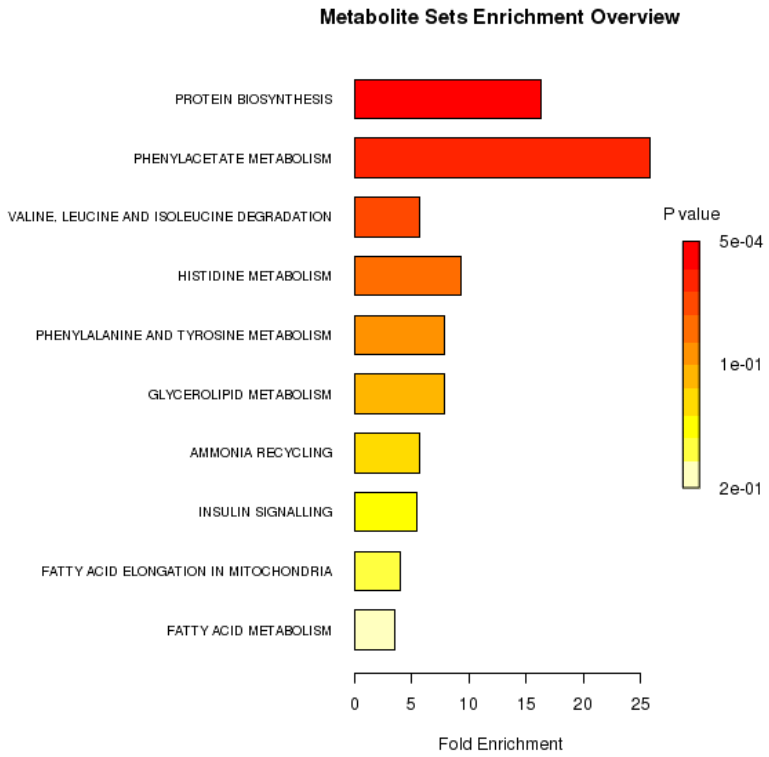

Figure 20: Metabolite set enrichment analysis results.

Protein biosynthesis was the only pathway to remain significant following FDR correction (FDR-adjusted P=4.33x10^{-3}, Figure 20). In total, ten pathways were found to be nominally enriched in this dataset (P <0.05, Figure 20). In addition to the protein biosynthesis pathway, these were phenylacetate metabolism, valine, leucine and

isoleucine degradation, histidine metabolism , phenylalanine and tyrosine metabolism, glycerolipid metabolism, ammonia recycling, insulin signalling, fatty acid elongation in mitochondria and fatty acid metabolism. The metabolites enriched in the only significant pathway, protein biosynthesis, were phenylalanine, isoleucine and histidine and these results suggest that changes in the human gut microbiota are linked to obesity and may influence protein metabolism.

6.2.4 *Microbial Metabolites and Kidney Function*

Chronic kidney disease (CKD) is characterised as the loss of renal function over a prolonged period of time, often months or years, however to be distinguished from acute renal failure, the loss of function must have been experienced for longer than 3 months. Symptoms include fatigue, increased blood pressure, accumulation of urea and potassium and anaemia, and the test for CKD measures blood creatinine. Glomerular filtration rate (eGFR) is the rate at which waste products are filtered from the blood. High levels of creatinine indicates a reduction in eGFR and consequently a reduction in kidney function. Three metabolites, indoxyl sulfate, p-cresol sulfate and phenylacetylglutamine, were found to be inversely correlated with eGFR in the TwinsUK cohort [Barrios et al., 2015]. Indoxyl sulfate and p-cresol sulfate are both uremic toxins and all three metabolites are microbially-derived, providing some evidence to suggest that the human microbiota may influence CKD. I was therefore interested to determine the associations between these three metabolites and the human microbiota in twins, as well as characterising the associations between chronic kidney disease and the gut microbiota.

To begin with, I considered the microbiota-metabolite results for indoxyl sulfate, p-cresol sulfate and phenylacetylglutamine

in the overall microbiome-metabolite study results outlined in Section 6.2.1. 52 16S OTUs were significantly associated with phenylacetylglutamine and 3 with indoxyl sulfate at an FDR threshold of 5% (Figure 21). The top association was between the heritable microbe family *Christensenellaceae* and phenylacetylglutamine ($P=3.54x10^{-07}$). *Christensenellaceae* was not significantly associated with either p-cresol sulfate or indoxyl sulfate however, at any level of significance.

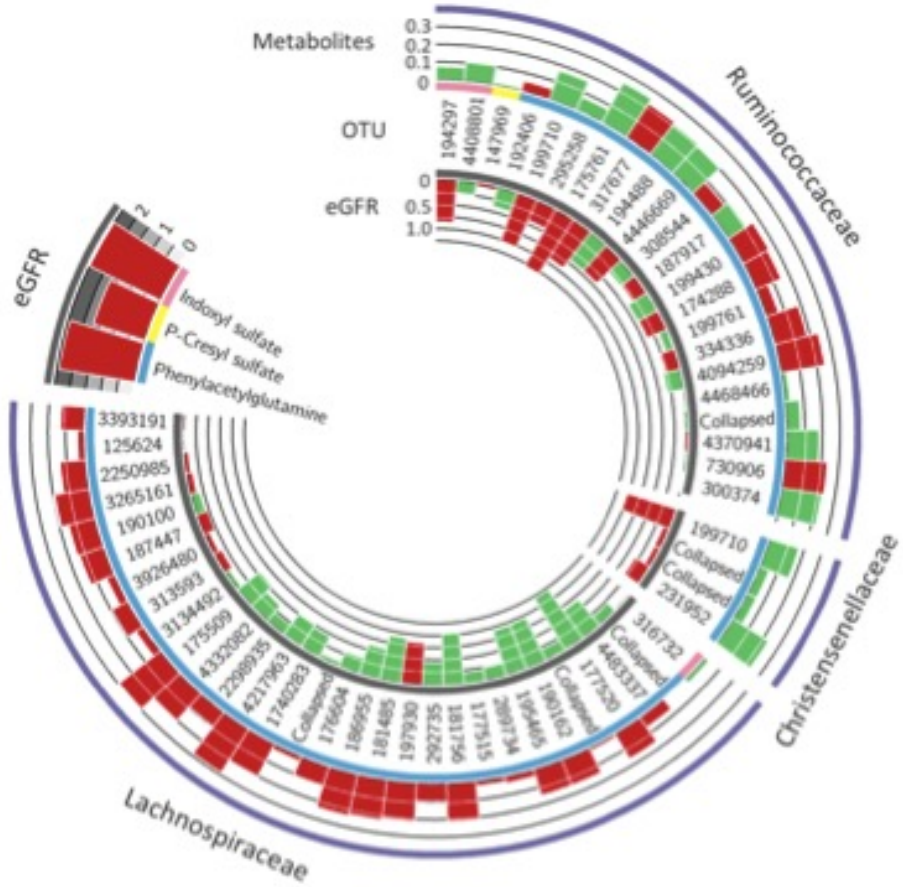

Figure 21: . Interactions between eGFR, indoxyl sulfate, phenylacetyleglutamine, p-cresol sulfate and the human microbiome. Red indicates an inverse association while green indicates a positive association. Column heights represent the beta coefficient.Adapted from Barrios and Beaumont et al, *PLoS ONE* 2015 [Barrios et al., 2015]

The majority of the OTUs and collapsed taxonomies that were significantly associated with these three metabolites, belonged to the *Ruminococcaceae* and *Lachnospiraceae* families, unsurprising given

that these two families are the most abundant in the human gut microbiota.

A subset of 854 healthy subjects had data on CKD outcome (a binary trait denoting disease status), eGFR and 16S microbial profiles. Of these 854, 55 (6.4%) individuals had impaired eGFR (<60 eGFR mL/min/1.73m^2). To test the association between the human gut microbiota, a linear mixed effects model was used (as described in (Chapter 2), accounting for twin structure, zygosity and sex. No CKD outcome associations surpassed FDR 5%. 46 associations were nominally significant between 16S data and CKD outcome (Table 18). The top association was negative with an unknown *Lachnospiraceae* OTU (Greengenes 352529, P=0.0052).

OTU	Family	Genus	P Value	Std Err	Estimate
352529	Lachnospiraceae	Unknown	0.00517	0.02881	-0.08080
4447950	Bacteroidaceae	Unknown	0.00794	0.01136	-0.03033
Collapsed	Lachnospiraceae	Lachnobacterium	0.00872	0.01554	0.04096
199677	Unknown (Clostridiales)	Unknown	0.00910	0.03467	-0.09085
3943186	Lachnospiraceae	Lachnobacterium	0.00974	0.01555	0.04040
151870	Erysipelotrichaceae	Coprobacillus	0.01055	0.02357	-0.06043
4469007	Unknown (Clostridiales)	Unknown	0.01179	0.00843	0.02132
181167	Lachnospiraceae	Dorea	0.01283	0.02176	0.05435
189937	Ruminococcaceae	Faecalibacterium prausnitzii	0.01291	0.03402	0.08477
2724175	Lachnospiraceae	Unknown	0.01295	0.01539	-0.03838
Collapsed	Clostridiaceae		0.01341	0.00813	0.02019
Collapsed	Clostridiaceae	Clostridium	0.01431	0.00536	0.01319
4202174	Clostridiaceae	Unknown	0.01478	0.01065	0.02623
Collapsed	Clostridiaceae	SMB53	0.01596	0.01068	0.02593
Collapsed	Odoribacteriaceae		0.01658	0.00440	0.01063
305318	Lachnospiraceae	Unknown	0.01741	0.01883	0.04517
193477	Lachnospiraceae	Unknown	0.01943	0.03084	0.07239
2740950	Lachnospiraceae	Coprococcus	0.01963	0.01819	0.04281
184465	Lachnospiraceae	Unknown	0.02134	0.03534	-0.08187
Collapsed	Clostridiaceae	Unknown	0.02164	0.00842	0.01944
4443846	Lachnospiraceae	Unknown	0.02370	0.02012	-0.04562
2123717	Lachnospiraceae	Unknown	0.02527	0.03413	0.07657
3275562	Lachnospiraceae	Unknown	0.02647	0.00782	0.01739
4435400	Lachnospiraceae	Unknown	0.02706	0.02095	-0.04649
182054	Lachnospiraceae	Coprococcus	0.02924	0.02179	-0.04774

Table 18: Top 25 significant 16S associations with CKD outcome.

When analyses were repeated using estimated glomerular flow rate (eGFR) as the phenotype, 23 associations had nominal significant between 16S profiles and eGFR (Table 19). The unknown *Ru-*

minococcaceae OTU, 198221, was the top hit (P=0.01). No associations surpass FDR 5%.

OTU	Family	Genus	P Value	Std Err	Estimate
198221	Ruminococcaceae	Unknown	0.01062	0.04417	-0.11377
4447950	Bacteroidaceae	Bacteroides	0.01610	0.02038	0.04926
186022	Lachnospiraceae	Blautia	0.01804	0.05092	0.12062
266274	Ruminococcaceae	Unknown	0.02016	0.03004	-0.06994
Collapsed	Lachnospiraceae	Coprococcus	0.02168	0.00709	0.01631
2256425	Christensenellaceae	Unknown	0.03142	0.04295	-0.09256
Collapsed	Peptococcaceae	Unknown	0.03312	0.04347	0.09321
Collapsed	Clostridiaceae	SMB53	0.03316	0.01927	-0.04121
4448492	Lachnospiraceae	Unknown	0.03323	0.02229	0.04760
182054	Lachnospiraceae	Coprococcus	0.03361	0.03922	0.08385
180721	Ruminococcaceae	Unknown	0.03381	0.03682	-0.07857
179512	Clostridiaceae	SMB53	0.03514	0.06540	-0.13811
288810	Ruminococcaceae	Unknown	0.03935	0.02732	-0.05646
191779	Lachnospiraceae	Blautia	0.03966	0.05396	-0.11159
187504	Ruminococcaceae	Unknown	0.04039	0.01583	-0.03249
185814	Ruminococcaceae	Unknown	0.04187	0.02457	-0.05017
Collapsed	Burkholderiales (class)		0.04218	0.00937	0.01922
782953	Enterobacteriaceae	Unknown	0.04240	0.01444	-0.02963
4454531	Enterobacteriaceae	Unknown	0.04481	0.03552	-0.07178
4427290	Ruminococcaceae	Unknown	0.04573	0.02131	0.04268
Collapsed	Betaproteobacteria (class)		0.04579	0.00959	0.01935
Collapsed	Enterbacteriaceae		0.04839	0.01378	-0.02760

Table 19: Nominally significant associations between 16S profiles and eGFR.

While the direct associations between the microbiota and CKD described here are weak, there are some significant associations between the microbiota and serum metabolites that have been identified as potential biomarkers for CKD. This provides supportive evidence for the role of the human gut microbiota in CKD progression. These data were publish in Barrios and Beaumont (2015) *PLoS ONE* and the manuscript can be found in Appendix B, Section B.2.

6.2.5 *Faecal Metabolomics*

All the results described so far in this chapter have concerned blood serum and plasma metabolites. The following section will describe preliminary faecal metabolomic analyses. The data is described fully

in Chapter 2. Faecal metabolomics can be used to measure microbial metabolic activity in the large intestine, however not all microbial metabolites are excreted in faeces and so this method will not produce a comprehensive and accurate picture of gut microbiota metabolism. P-cresol for example, while produced in the gut via microbial fermentation of tyrosine, phenylalanine and tryptophan, is readily absorbed by the mucosal cells in the colon and excreted in urine [Roediger and Babidge, 1997].

Despite their potential usefulness, faecal metabolomics studies remain few in number. It is unclear why this is the case, however one study suggests it may be due to a lack of characterisation of the 'normal' faecal metabonome [Saric et al., 2008], and so made efforts to characterise exactly that in mice, rats and humans. They determined that the faecal metabolomic profile is not stable over time and has unique differences across species that may present issues if using mice or rats as a model organism for humans [Saric et al., 2008].

I was first interested to explore faecal metabolite associations with obesity and then to determine human gut microbiome associations with faecal metabolites. In this pilot dataset of faecal metabolites, there are a total of 60 individuals, all of which are MZ twins. A subset of 49 individuals had adiposity data.

6.2.5.1 *Faecal Metabolomics and Obesity*

After normalisation, QC and imputation, I first wanted to determine faecal metabolite associations with obesity. Using a linear mixed effects regression, I considered adiposity (either visceral fat, % Trunk Fat or BMI) as the response variable, metabolite as the predictor and BMI (except in the case of BMI as the test variable), age and family structure as covariates (family structure was considered as a random effect.).

When all 3 adiposity phenotypes are considered no associations surpass Bonferonni. 297 associations surpass a nominal significance threshold of P<0.05. The top association was between VFM and N-acetylleucine (P=0.00027), followed closely by an association between VFM and N-6-trimethyllysine (P=0.00028, Table 20. Full table can be found in Appendix A, Table A5).

Metabolite	Pheno	P Value	Std Err	Estimate	FDR Q Value
N-acetylisoleucine	VFM	0.0003	0.0826	0.3379	0.2469
N-6-trimethyllysine	VFM	0.0003	0.0830	0.3238	0.2469
X-21742	BMI	0.0003	0.1726	-0.7134	0.2469
pseudouridine	VFM	0.0008	0.0905	0.3232	0.2790
S-adenosylmethionine (SAM)	VFM	0.0009	0.0943	0.3335	0.2813
N-acetyltyrosine	VFM	0.0010	0.0896	0.3130	0.2841
N-acetyltyrosine	% Trunk Fat	0.0011	0.0941	0.3270	0.2861
X-12095	% Trunk Fat	0.0012	0.0947	0.3293	0.2873
X-13696	% Trunk Fat	0.0012	0.0990	-0.3438	0.2875
X-12700	VFM	0.0012	0.0814	0.2777	0.2882
X-11612	VFM	0.0013	0.0870	0.2961	0.2891
2-pentanamido-3-phenylpropanoic acid	VFM	0.0016	0.0968	0.3259	0.2994
2-aminoheptanoate	VFM	0.0019	0.0824	0.2697	0.3065
X-17438	VFM	0.0019	0.0867	-0.2842	0.3078
2-hydroxyoctanoate	% Trunk Fat	0.0026	0.0897	0.3120	0.3198
5-methylthioadenosine (MTA)	VFM	0.0027	0.0876	0.2755	0.3209
uracil	VFM	0.0029	0.0842	0.2635	0.3235
5-methylthioadenosine (MTA)	% Trunk Fat	0.0030	0.0917	0.2879	0.3241
lysylleucine	VFM	0.0030	0.0890	0.2884	0.3248
5-methyluridine (ribothymidine)	VFM	0.0032	0.0970	0.3038	0.3262

Table 20: Top 20 associations between adiposity and faecal metabolites.

When I performed MSEA on nominally significant metabolites associated with visceral fat, the protein biosynthesis pathway was significantly enriched ($P = 5.83 \times 10^{-11}$) and remains so after FDR correction ($Q = 4.67 \times 10^{-9}$). The small sample size and observational nature of this study prevents any further determination as to how this pathway may be affected.

6.2.5.2 *Faecal Metabolomics and the Human Gut Microbiota*

I was next interested in 16S associations with faecal metabolites. Using a linear model as performed in Section 6.2.1, I tested the association between OTUs and faecal metabolites. 22 associations surpassed

Bonferonni threshold of p=5.4x10^{-8}. This equated to 21 unique microbial units and 19 unique metabolites. The top association was between isopalmitic acid and an unknown *Lachnospiraceae* OTU (Greengenes OTU 311820, $P = 1.08 \times 10^{-9}$, Table 21) and additionally, a large number of the top hits were composed of *Lachnospiraceae* and *Ruminococcaceae* associations but only 5 of the Bonferonni-significant OTUs had associations with Bonferonni-significant OTUs in adiposity (Section 5.2.3)

OTU	Parent Taxonomy	Taxonomy	Metabolite	P Value	Beta	SE
311820	Lachnospiraceae	Unknown	isopalmitic acid	1.08E-09	-2.0269	0.1814
302160	Ruminococcaceae	Unknown	X-22062	1.45E-09	-1.0344	0.0935
197364	Clostridiales (Order)	Unknown	13-methylmyristic acid	2.52E-09	-0.4932	0.0121
178478	Bacteroidaceae	Bacteroides unknown	aspartylleucine	3.27E-09	-0.9131	0.0385
193915	Ruminococcaceae	Unknown	inositol-1-phosphate (I1P)	5.21E-09	2.2439	0.1416
187180	Ruminococcaceae	Unknown	valylserine	1.06E-08	-0.8753	0.0246
340711	Ruminococcaceae	Unknown	X-17698	1.72E-08	2.3798	0.1938
182911	Ruminococcaceae	Unknown	leucyltyrosine	1.75E-08	-0.6464	0.066
2875735	Bacteroidaceae	Bacteroides unknown	xanthosine	2.42E-08	1.3294	0.0291
332929	Ruminococcaceae	Unknown	threonine	2.60E-08	1.5405	0.0324
182911	Ruminococcaceae	Unknown	threonylarginine	2.82E-08	-0.6407	0.0644
537219	Clostridiales (Order)	Unknown	lysine	2.92E-08	-0.9477	0.0834
179744	Ruminococcaceae	Oscillospira (unknown species)	X...17303	3.02E-08	2.2104	0.2442
3195723	Ruminococcaceae	Oscillospira (unknown species)	X...14697	3.21E-08	-1.1660	0.0635
Collapsed	Firmicutes (Phylum)	Bacilli	N.methylpipecolate	3.55E-08	-0.6708	0.0282
535955	Clostridiaceae (	SMB53 (Unknown species)	deoxycarnitine	4.33E-08	0.7161	0.0815
198422	Clostridiales (Order)	Unknown	N.acetylthreonine	4.35E-08	0.9187	0.0484
Collapsed	Tenericutes (Phylum		beta.hydroxyisovalerate	5.01E-08	-0.3577	0.0210
Collapsed	RF3 (Class)	ML615J.28	beta.hydroxyisovalerate	5.01E-08	-0.3577	0.0210
Collapsed	ML615J.28 (Order)	Unknown	beta.hydroxyisovalerate	5.01E-08	-0.3577	0.0210
Collapsed	ML615J.28 (Order)	Unknown	beta.hydroxyisovalerate	5.01E-08	-0.3577	0.0210
174439	Ruminococcaceae	Faecalibacterium prausnitzii	X...20172	5.30E-08	-1.1966	0.0536

Table 21: Bonferonni-significant associations between 16S microbiome and faecal metabolites.

In summary, the faecal metabolite pilot provided promising results that could encourage full-scale analysis. While the sample was fairly small at 60 samples, this is in fact the largest study of its kind. A large number of associations were observed between the human microbiota and faecal metabolites, as expected, however no adiposity associations surpassed multiple testing thresholds. This may be simply due to a small sample size and should be readdressed in a larger study.

6.3 CONCLUSION

In this chapter I was primarily interested in the relationship between the human gut microbiota and metabolic end-products. Microbiota associations with metabolites can be quite strong, as demonstrated by some of the high p-values seen in this Chapter, as well as the sheer number of associations that surpass an FDR threshold of 1%. This makes finding meaningful, biological stories somewhat difficult to tease apart and taking a narrower view using candidate metabolites can produce clearer insights while awaiting larger samples, replication and better tools to fully understand the bigger picture. I demonstrate this in Section 6.2.4 with kidney function and metabolites. Faecal metabolites can provide an insight into the metabolic processes taking place within the gut, and in the case of metabolites that are associated with visceral fat, protein synthesis may be impacted in instances of altered visceral fat mass.

7

DISCUSSION

This body of work is one of the largest population-based microbiome studies to date in one of the most well-studied cohorts. It has explored heritability of adiposity and the human gut microbiome, determined that the human gut microbiome is strongly associated with adiposity as well as specific environmental exposures, investigated microbiota changes associated with renal function decline and explored the effect of *Christensenella* in the host.

Using the ICC method, I determined heritability of 16S microbial profiles in a set of over 900 twins. This was later confirmed by colleagues at Cornell University using structural equation modelling and the ACE model [Goodrich et al., 2014]. This is the first heritability study that has been performed on 16S data, however re-analysis of the publicly available Yatsunenko et al. [2012] and Turnbaugh and Gordon [2009] datasets yielded similar results thus providing validation. *Christensenella* was the most heritable microbe in the TwinsUK dataset while also being highly heritable in the Yatsunenko dataset. With the observation that levels of *Christensenella* were higher in lean twins vs obese twins, our collaborators further explored the role of *Christensenella* in obesity. When mice inoculated with *Christensenella* were fed high fat diets, the mice that received *Christensenella* displayed attenuated weight gain. This hints at the possibility that *Christensenella* may play a preventative role in the progression of obesity however little is known and published on this potentially important microbe.

Given the apparent importance of *Christensenella* I explored its association with additional phenotypes in the TwinsUK cohort in

Chapter 4. I first explored its association with additional adiposity phenotypes, and found strong associations with visceral fat and % trunk fat, as well as moderate associations with BMI, confirming the role of this microbe in obesity.

As explained in Chapter 4, *Christensenella* is an obligate anaerobe with the ability to produce butyrate and acetate. While butyrate is considered to be a beneficial metabolite for the host due to its anti-inflammatory properties [Meijer et al., 2010], increased SCFA, particularly proprionate, have been linked with obesity in a few small to medium-sized studies [Payne et al., 2011, Schwiertz et al., 2010] likely due to increased energy harvest from food or increased carbohydrate consumption [Duncan et al., 2008]. The apparent protective association between the butyrate-producing *Christensenella* and obesity is in contrast to this, however *Christensenella* is not responsible for the production of all the SCFA in the human host. Potentially, the benefits of increased butyrate may overshadow the effects of other SCFA. Unfortunately, the TwinsUK cohort does not possess measures of SCFA at present and so I was unable to confirm or deny the link between SCFA and obesity in a large sample. *Christensenella* also co-occurs with members of the methanogen-producing Archaea [Goodrich et al., 2014], potentially forming a syntrophic relationship. The methanogens form other syntrophic relationships however, and these have been linked with increased host adiposity, both in mice [Samuel and Gordon, 2006] and in humans [Zhang et al., 2009]. In mice, *Methanobrevibacter smithii* was able to modulate the activity of *Bacteroides thetaiotaomicron* so that it produced more acetate and less butyrate [Samuel and Gordon, 2006]. Acetate contributes to lipid synthesis and cholesterol production [Wolever et al., 1995] which would explain the increased adiposity observed in mice possessing both *M. smithii* and *B. thetaiotaomicron*. I hypothesise that it may not be the

presence of *Christensenella* that is important for decreased host adiposity, but given that *M. smithii* is widely shared and present in 64% of this twin sample, but a possible syntrophic relationship between the methanogens and *Christensenella* that may stimulate butyrate production instead of acetate and therefore contribute to minimising inflammation and providing energy for colonic epithelial cells, instead of lipid metabolism.

Indeed, my findings do seem to support the hypothesis that *Christensenella* may be involved with inflammation. *Christensenella* was strongly and negatively associated with visceral fat (Section 4.2.1), a metabolically significant adipose tissue, capable of secreting the pro-inflammatory cytokine, IL-6. The direction of association indicates that in instances of increased visceral fat, there tends to be a reduction of *Christensenella*. Furthermore, *Christensenella* tends to be reduced when white blood cell counts such as monocytes and neutrophils are raised (Section 4.2.2). Candidate gene analysis of known human immune regions revealed a number of suggestive associations between *Christensenella* and *ENTPD1*, the gene that encodes the regulatory T-cell (Treg) CD39 that is important in suppressing inflammatory responses by degrading ATP to AMP [Antonioli et al., 2013] (Section 4.2.3.2). Given the anti-inflammatory effect of *ENTPD1* and its prominent location upon monocytes and neutrophils, both of which associated with *Christensenella*, it presents an interesting association worthy of further follow-up. Despite all these findings being observational in nature and not indicative of causality, they do support the hypothesis that *Christensenella* is inversely linked with inflammation.

As *Christensenella* may be beneficial for health it is natural to question how more *Christensenella* can be fostered in the host microbiota. Given that *Christensenella* produces SCFA it is prudent to assume that *Christensenella* would therefore thrive on a high fibre

diet. Indeed, the presence of *Christensenella* in the aquatic mammal, dugong, would seemingly support this theory. D. dugon is a marine hind-gut fermenter, requiring the colonic microbiome to break down eelgrass, more specifically in the case of this dugong, *Zostera marina*. *Zostera* species of eelgrass are more fibrous than other species, with one species, *Zostera capricornii*, which has a mean neutral detergent fibre (NDF) level of 63% of dry mass [Marsh et al., 2012]. As terrestrial grasses typically have 50-70% NDF, this places *Zostera* species at quite a high level of fibre content.

The primary focus of my thesis was on exploration of gut microbiome profiles in obesity as described in Chapter 5. While this is an area of research that has been studied in depth, previous studies have used BMI as a measure of adiposity. BMI is an index of overall body mass and while an adequate measure the majority of the time, it can result in children or sportsmen being falsely considered as obese. This has been demonstrated in rugby players where their BMI's were ~29 or above and so considered obese, yet their weight was due to muscle mass. These individuals also had an higher microbial diversity than both high and low BMI controls, potentially due to their diet and high amount of exercise [Clarke et al., 2014]. A number of adiposity measures were used in this thesis, including BMI, visceral fat and android:gynoid ratio. Each of these measures a different aspect of obesity; BMI measures overall body mass, irrespective of fat and lean mass, android:gynoid ratio adjusts truncal or android adiposity for gynoid adiposity as the latter is held to be beneficial for health [Snijder et al., 2005] and visceral fat is primarily concerned with the deep adipose tissue that surround organs and is strongly linked with disease. Due to the metabolic consequences of increased visceral fat, chiefly insulin resistance and inflammation [Fontana et al., 2007], measures of visceral fat are more clinically rele-

vant than BMI despite being harder to obtain. I observed that associations between visceral fat and the microbiome are stronger than those with BMI and in addition, the families of microbes that associate with BMI, Ruminococcaceae and Lachnospiraceae, associate with visceral fat estimates also. Yet there is no consensus with direction of effect within these families. Some genera are more abundant in obese individuals while others are less abundant and this highlights the importance of looking at finer-grained taxonomic resolution when considering microbiome alterations in disease. A number of the microbes that are associated with visceral fat are also associated with altered serum metabolite profiles, particularly palmitate, phenylalanine, histidine and isoleucine, all of which have previously been linked with metabolic disorders (Table 22).

Metabolite	Cardiometabolic Disease Link
Palmitate	Induces insulin resistance[Thrush et al., 2008]
Phenylalanine	Highly associated with diabetes[Wang et al., 2011]
Histidine	Ameliorates inflammation in obese women and rats[Sun et al., 2014]
Isoleucine	Prevents accumulation of triglycerides in mice[Nishimura et al., 2010]

Table 22: Links between significant metabolites and cardiometabolic disease.

Palmitate has been linked with insulin-resistance in obesity [Gao et al., 2010, Thrush et al., 2008] however given that it is the most abundant fatty-acid in food, it is entirely possible that these alterations are driven by diet. Enrichment analysis revealed that the protein biosynthesis pathway is altered with the alteration of these metabolites. Protein biosynthesis is concerned with the creation of lean mass, primarily muscle mass, and it is thought to be impaired in obese individuals, leading to higher fat mass [Guillet et al., 2011]. That microbes associated with obesity also associated with metabolites involved with protein synthesis is perhaps indicative of the intricate network between the human gut microbiome and metabolic processes.

Environmental effects are important to consider when analysing microbiome data. I explored the impact of three specific environmental factors, smoking, diet and method of delivery, on the gut microbiome in twins in Chapter 3. Diet is presumed to be the largest factor affecting the human microbiome and numerous studies have shown the effects of a wide range of diets on the gut microbiota. In this dataset, I found few significant associations and there are a number of reasons why this might be. The first and most obvious potential reason is the time difference between the dietary data collection and the faecal sample collection. The FFQ data was collected prior to 2007 while the faecal collections began in 2010. Another potential reason is that in a large sample size, the effect of a fatty diet may no longer present, but the wealth of evidence in animal studies suggests this may not be the reason for differing results. However it is worth remembering that results in animal models may not necessarily translate into human subjects. Lack of dietary evidence in this cohort may be due to sampling. With any questionnaire there is a risk of participant bias due to the unwillingness of a subject to admit to certain habits. This can result in the alteration of answers to make the subject appear, for example, to be a healthier eater than they actually are. Nor does the subject have to lie about their dietary habits to create noise in the data; simply overestimating or underestimating consumption of certain foods, however consciously or sub-consciously, will affect dietary data outcomes. The TwinsUK cohort has had some success in using its dietary data to find meaningful associations however, suggesting that the dataset is reliable. Heritable food patterns were discovered accounting for 5 broad dietary patterns including traditional english, dieting, fruit and veg, low meat and high alcohol [Teucher et al., 2007]. In addition, preferences for garlic and coffee were found to be highly heritable [Teucher et al., 2007]. Another study found

that wine and traditional english diet both had effects on bone mineral density in the same cohort [Fairweather-Tait et al., 2011]. FFQ questionnaires may simply be an inaccurate way to measure and diet intervention studies may be more effective. This involves placing participants on specific diets and measuring the effect on the microbiome. In this way, dietary intakes will be controlled and study groups can be selected for comparison.

Method of delivery has been reported to affect the gut microbiome until as late as 7 years of age [Salminen et al., 2004]. In this sample, there were no strong associations between method of delivery and the gut microbiome. This could be due to the low incidence of caesarean section in the sample (3.9%)and it is unlikely there was reasonable power to detect the effect method of delivery might have on the microbiome, especially given the current age of the subjects. Furthermore, infants born by caesarean section may be premature and possess low birth weight which may confound the results.

The effect that tobacco smoking has on the human gut microbiome is relatively unstudied. In one study smoking was found to affect the gut microbiome in such a way that it might influence obesity [Biedermann et al., 2013]. Indeed it does seem that microbial diversity is unaffected by smoking as shown in Chapter 3, in agreement with Biederman et al (2013). However, I observed a negative association between the beneficial microbe, *B. animalis* and smoking status which may be of interest. Smokers are at a higher risk of developing the inflammatory disorder, Crohn's disease [Cottone et al., 1994] but not ulcerative colitis [Lindberg et al., 1988]. *B. animalis* spp. *lactis* has been found to reduce inflammatory markers in a mouse model of colitis [Philippe et al., 2011] however *B. animalis* has also been found to induce inflammation in IL-10 deficient gnotobiotic mice [Moran et al., 2009]. *B. animalis* can refer to one of two subspecies; *B. animalis* spp.

lactis or *B. animalis* spp. *animalis*. The different effects of these subspecies may account for the differences observed in these two studies and so if *B. animalis* was to be considered as a probiotic treatment for gastrointestinal trouble in smokers, the subspecies would have to be selected carefully. A probiotic intervention in humans using these 2 subspecies of *B. animalis* would be a good way to replicate the difference in effect of these two bacteria, and may also result in a viable candidate to treat gastrointestinal upset in smokers.

As presented in Chapter 3, and discussed above, the gut microbiome has heritable components and has been shown here for the first time, particularly species and genera within the Firmicutes phylum [Goodrich et al., 2014]. While the sample size of this dataset is relatively large for a microbiome study, this is considerably small sample for GWAS of human traits, where modest genetic effects require much larger samples to ensure good power to detect associations. Furthermore, the microbiome heritability estimates from the ACE model in Chapter 3 were not strong, suggesting that we expect modest and moderate host genetic effects on the human gut microbiome. Instead I used a targeted candidate approach to perform genetic studies. Using previously published obesity genetic variants from the NHGRI-EBI GWAS Catalog I performed a candidate gene analysis with the gut microbiome, presented in Chapter 4 and Chapter 5. *RPTOR* was the top association and encodes the protein raptor, which regulates the activation of IRS-1 and adipogenesis in response to nutrient and insulin levels through binding of mTOR [Tzatsos and Kandror, 2006]. Chronic activation of the mTOR pathway has been shown to lead to insulin resistance and obesity [Patti and Kahn, 2004]. The results show that Lachnospiraceae is less abundant when there is an increase in insulin and adipose tissue, but the exact mechanism linking *RPTOR* with Lachnospiraceae remains unknown. Causal models did not

help clarify the relationship. It is possible that *RPTOR* has epigenetic or expression effects on colonic epithelium (as suggested by gene expression profiles from the GTEX consortium dataset), which may lead to differences in the gut microbiota.

Metabolomics provide a snapshot into the metabolic processess that take place within the host. A number of metabolites that can be measured in human blood and faeces are created by the host microbiota, thus metabolomics can also provide a snapshot of microbiota metabolic processes. In Chapter 6 I explored the associations between the human gut microbiota and serum, plasma and faecal metabolites. Serum/plasma metabolites that have been linked with host health were found in this project to be associated with members of the gut microbiome in twins. In particular, the cardio-metabolically important palmitate formed a number of associations. Palmitate has significant implications on host health, yet is also the most abundant fatty acid found, which may explain the number of associations. A number of studies have shown palmitate [Thrush et al., 2008, Hoppa et al., 2009, Feng et al., 2012] is responsible for inhibiting insulin secretion and the adverse effect of trans fatty acids is well documented [Willett, 2012], making palmitate an important contributory factor to diabetes and cardiovascular disease. How the gut microbiome links to palmitate is unclear. One metabolite that was significantly associated with members of the human gut microbiome was phenylacetylglutamine. Phenylacetylglutamine is in fact a microbial co-metabolite [Wijeyesekera et al., 2012] and has been linked with end stage renal failure in type 2 diabetic patients [Niewczas et al., 2014]. To this end, I was interested to explore the association between the human gut microbiota and renal function.

In collaboration with my colleague, Dr Clara Barrios, we explored the association between renal function, as measured by esti-

mated glomerular filtration rate, serum/plasma metabolites and the human microbiome. Firstly, 3 metabolites were identified as being significantly and negatively associated with renal function: *p*-cresol sulfate, phenylacetyleglutamine and indoxyl sulfate. All three of these metabolites are microbial co-metabolites which proposed the hypothesis that the human gut microbiome may be implicated in the progression of chronic kidney disease. Christensenellaceae, Ruminococcaceae and Lachnospiraceae had significant associations with phenylacetylglutamine and indoxyl sulfate. When associations between the gut microbiome and eGFR were tested, there were few significant hits and none were particularly strong. This can be explained by the predominantly healthy nature of this cohort. Strong associations with the microbial metabolites identified as potential CKD biomarkers however suggest a potential link between CKD and the human microbiome that should be explored further in a disease-specific cohort. These results have been published in PLoS ONE [Barrios et al., 2015].

Finally I performed a small-scale pilot analysis, comparing faecal metabolites to adiposity and 16S microbiome profiles. A number of metabolites were nominally significantly associated with adiposity, although these associations did not surpass a multiple testing threshold. Metabolite set enrichment analysis revealed that a number of the significant metabolites were involved in the protein biosynthesis pathway. Potentially, protein biosynthesis may be impacted by alterations in the colonic microbiome, mediated by metabolites, that affect obesity. Microbial associations with faecal metabolites were much stronger, resulting in many more associations than with adiposity. The top association was with isopalmitic acid, a related metabolite to the peak metabolite, palmitate, in the serm/plasma metabolite analysis. The sample size for this study is small, and more samples are required to detect associations with adiposity, however associations

with the 16S profiles are strong, confirming the strength of this technique in determining a snapshot of microbial metabolic activity in the gut. There are few faecal metabolomics studies published and as such, it is unclear what is considered a "normal" faecal metabolome [Saric et al., 2008]. While this may be a powerful technique, it requires further work to establish baseline measurements before comparisons can be performed.

In summary, I have contributed to the discovery of genetic influence on microbial abundance and found some key BMI associations. In addition, I have characterised the microbiome differences in obesity using a novel adiposity measure- visceral fat, performed exploratory analyses to characterise *Christensenella* further, developed hypotheses about its role in obesity and inflammation and found evidence to support such a role. I have performed candidate gene analyses to discover host genes that may impact microbial communities and used metabolite profiles to determine changes in early renal decline. I have also explored known environmental factors in this dataset, while also looking at smoking, a relatively unknown factor affecting the microbiome. This work has brought the effect of genetics to the attention of the microbiome research community and stressed the importance of considering host genetic effect of microbial communities. It has further confirmed the need to take obesity into account when performing analyses or participant selection.

To further this work I would move from observational studies and begin to perform functional analyses. In order to define the role of *Christensenella* in obesity I would perform similar mouse experiments to those in Goodrich et al (2014), this time monitoring cytokines in the blood prior to inoculation with *Christensenella* and then at various time points following inoculation. Transcriptomic analysis on the *Christensenella* microbes would also be ideal, in order to

see which genes are expressed to gain an idea as to what this important microbe might be doing to influence obesity of inflammation. In fact, a meta-transcriptomic analysis on the entire human microbiome would be an interesting study to perform in order to understand functional alterations in obesity. Dietary interventions would also be an option to further a number of the studies performed in this thesis. Firstly, one study could be aimed at determining the ideal prebiotics that would promote *Christensenella* growth, while a second study could focus on probiotic supplementation to improve the gastrointestinal health of smokers. Collection of more faecal and DNA samples in order to perform a full genome-wide association study would also be a priority, although many thousands would have to be collected. This would help to find host genes that may affect microbial composition. In order to address questions regarding butyrate, I would measure SCFAs and determine associations with obesity and Christensenella. This may lead to further mouse models for causal investigation.

Part III

APPENDIX

A

APPENDIX A: RESULTS TABLES

This appendix lists expanded results tables for the thesis.

Table A1: ACE and ICC heritability estimates of all microbial units in the TwinsUK cohort.

OTU	Parent Taxonomy (Level)	Taxonomy (Level)	ACE_A	ACE_C	ACE_E	ICC_A	ICC_C	ICC_E
Collapsed	Christensenellaceae (Family)	Unknown Genus	0.388	1.56E-17	0.612	0.604	-0.194	0.591
Collapsed	Clostridiales (Order)	Christensenellaceae (Family)	0.382	4.20E-13	0.618	0.609	-0.205	0.595
176318	Christensenellaceae (Family)	Unknown Genus and Species	0.355	1.95E-13	0.645	0.542	-0.175	0.633
4347159	Bifidobacteriaceae (Family)	Bifidobacterium adolescentis	0.352	4.50E-15	0.648	0.675	-0.286	0.611
4468466	Ruminococcacaea (Family)	Unknown Genus and Species	0.344	2.85E-13	0.656	0.656	-0.257	0.601
Collapsed	Erysipelotrichaceae (Family)	cc_115 (Genus)	0.334	0	0.666	0.466	-0.125	0.660
Collapsed	Actinobacteria (Class)	Bifidobacteriales (Order)	0.328	2.03E-12	0.672	0.632	-0.260	0.628
Collapsed	Bifidobacteriales (Order)	Bifidobacteriaceae (Family)	0.328	2.03E-12	0.672	0.632	-0.260	0.628
Collapsed	Actinobacteria (Phylum)	Actinobacteria (Class)	0.321	5.50E-15	0.679	0.575	-0.223	0.649
819353	Ruminococcacaea (Family)	Unknown Genus and Species	0.315	1.74E-16	0.685	0.394	-0.091	0.697
Collapsed	Bifidobacteriaceae (Family)	Bifidobacterium (Genus)	0.312	0.0105	0.678	0.641	-0.269	0.628
Collapsed	Ruminococcaceae (Family)	Unknown Genus	0.300	1.82E-13	0.700	0.472	-0.151	0.679
292758	Clostridiales (Order)	Unknown Family, Genus and Species	0.297	5.29E-16	0.703	0.109	0.145	0.746
836693	SHA-98 (Order)	Unknown Family, Genus and Species	0.297	1.43E-15	0.703	0.469	-0.183	0.714
Collapsed	Clostridia (Class)	SHA.98 (Order)	0.297	1.43E-15	0.703	0.469	-0.183	0.714
Collapsed	SHA.98 (Order)	Unknown Family	0.297	1.43E-15	0.703	0.469	-0.183	0.714
Collapsed	SHA.98 (Order)	Unknown Family and Genus	0.297	1.43E-15	0.703	0.469	-0.183	0.714
289734	Lachnospiraceae (Family)	Unknown Genus and Species	0.296	8.85E-13	0.704	0.180	-0.072	0.892
4412540	Ruminococcacaea (Family)	Unknown Genus and Species	0.296	2.59E-15	0.704	0.493	-0.170	0.677
4421273	Ruminococcacaea (Family)	Oscillospira (unknown species)	0.294	2.57E-13	0.706	0.408	-0.108	0.700
1102370	Ruminococcacaea (Family)	Unknown Genus and Species	0.293	0.0365	0.671	0.250	0.032	0.717
540055	Ruminococcacaea (Family)	Unknown Genus and Species	0.292	0.0232	0.684	0.292	0.028	0.680
Collapsed	Clostridiales (Order)	Unknown Family	0.291	0	0.709	0.479	-0.201	0.721
Collapsed	Clostridiales (Order)	Unknown Family and Genus	0.291	0	0.709	0.479	-0.201	0.721
Collapsed	Clostridiales (Order)	Dehalobacteriaceae (Family)	0.291	6.63E-14	0.709	0.543	-0.214	0.671
4450214	Ruminococcacaea (Family)	Unknown Genus and Species	0.289	8.04E-13	0.711	0.393	-0.097	0.705
Collapsed	Euryarchaeota (Phylum)		0.288	0.0735	0.639	0.319	0.030	0.651
287790	Ruminococcacaea (Family)	Unknown Genus and Species	0.286	2.23E-12	0.714	0.132	-0.119	0.987

4370941	Ruminococcacaea (Family)	Unknown Genus and Species	0.285	9.03E-14	0.715	0.549	-0.218	0.669
Collapsed	Actinobacteria (Phylum)		0.285	5.01E-12	0.715	0.452	-0.160	0.708
4336939	Ruminococcacaea (Family)	Unknown Genus and Species	0.281	2.95E-13	0.719	0.720	-0.381	0.662
4300690	Ruminococcacaea (Family)	Unknown Genus and Species	0.279	0.0407	0.681	0.352	-0.020	0.669
Collapsed	Dehalobacteriaceae (Family)	Dehalobacterium (Genus)	0.279	9.10E-15	0.721	0.545	-0.221	0.676
183681	Ruminococcacaea (Family)	Unknown Genus and Species	0.275	0.0110	0.714	0.147	0.027	0.827
300374	Ruminococcacaea (Family)	Oscillospira (unknown species)	0.275	2.69E-27	0.725	0.138	-0.074	0.936
Collapsed	Peptostreptococcaceae (Family)	Unknown Genus	0.273	2.51E-13	0.727	0.386	-0.073	0.687
Collapsed	Christensenellaceae (Family)	Christensenella (Genus)	0.272	7.77E-15	0.728	0.300	-0.039	0.739
176062	Clostridiales (Order)	Unknown Family, Genus and Species	0.272	2.50E-13	0.728	0.190	0.040	0.770
Collapsed	Burkholderiales (Order)	Oxalobacteraceae (Family)	0.272	7.95E-14	0.728	0.593	-0.286	0.693
4153054	Ruminococcacaea (Family)	Unknown Genus and Species	0.270	1.39E-11	0.730	0.393	-0.135	0.742
4349261	Lachnospiraceae (Family)	Unknown Genus and Species	0.268	3.89E-13	0.732	0.424	-0.120	0.696
1504042	Ruminococcacaea (Family)	Oscillospira (unknown species)	0.266	0	0.734	-0.161	0.336	0.825
537219	Clostridiales (Order)	Unknown Family, Genus and Species	0.266	0	0.734	0.425	-0.127	0.703
Collapsed	Tenericutes (Phylum)		0.265	9.04E-14	0.735	0.555	-0.269	0.714
3195723	Ruminococcacaea (Family)	Oscillospira (unknown species)	0.262	3.23E-13	0.738	0.306	-0.053	0.746
3903651	Ruminococcacaea (Family)	Oscillospira (unknown species)	0.261	7.13E-12	0.739	0.263	-0.020	0.758
Collapsed	Actinomycetales (Order)	Actinomycetaceae (Family)	0.261	8.93E-14	0.739	0.328	-0.055	0.727
212532	Ruminococcacaea (Family)	Unknown Genus and Species	0.258	9.73E-14	0.742	0.127	0.014	0.860
199286	Clostridiales (Order)	Unknown Family, Genus and Species	0.256	1.84E-12	0.744	-0.017	0.083	0.934
177134	Ruminococcacaea (Family)	Unknown Genus and Species	0.253	2.03E-15	0.747	0.397	-0.146	0.749
Collapsed	Tenericutes (Phylum)	RF3 (Class)	0.253	3.37E-14	0.747	0.656	-0.328	0.673
Collapsed	RF3 (Class)	ML615J.28 (Order)	0.253	3.37E-14	0.747	0.656	-0.328	0.673
Collapsed	ML615J.28 (Order)	Unknown Family	0.253	3.37E-14	0.747	0.656	-0.328	0.673
Collapsed	ML615J.28 (Order)	Unknown Family and Genus	0.253	3.37E-14	0.747	0.656	-0.328	0.673
553080	Clostridiales (Order)	Unknown Family, Genus and Species	0.252	4.17E-12	0.748	0.306	-0.045	0.739
4364405	Ruminococcacaea (Family)	Unknown Genus and Species	0.252	5.84E-14	0.748	0.423	-0.150	0.727
Collapsed	Actinobacteria (Class)	Actinomycetales (Order)	0.251	0.0275	0.722	0.189	0.058	0.753
180136	Ruminococcacaea (Family)	Oscillospira (unknown species)	0.250	0.0141	0.736	0.352	-0.078	0.726
4453304	Lachnospiraceae (Family)	Unknown Genus and Species	0.246	2.50E-13	0.754	0.530	-0.242	0.712
174611	Ruminococcacaea (Family)	Faecalibacterium prausnitzii	0.246	0	0.754	0.324	-0.062	0.738

4469576	Lachnospiraceae (Family)	Unknown Genus and Species	0.244	0.0767	0.679	0.360	-0.013	0.653
72820	Bifidobacteriaceae (Family)	Bifidobacterium longum	0.244	0	0.756	0.473	-0.185	0.711
177567	Ruminococcacaea (Family)	Unknown Genus and Species	0.243	1.94E-13	0.757	0.303	-0.066	0.763
Collapsed	Tenericutes (Phylum)	Mollicutes (Class)	0.242	1.47E-14	0.758	0.399	-0.156	0.757
2943548	Ruminococcacaea (Family)	Ruminococcus (unknown species)	0.241	4.06E-11	0.759	0.393	-0.135	0.742
Collapsed	Lachnospiraceae (Family)	Blautia (Genus)	0.239	5.14E-13	0.761	0.307	-0.076	0.769
189828	Ruminococcacaea (Family)	Unknown Genus and Species	0.238	3.85E-13	0.762	0.158	-0.023	0.866
312882	Ruminococcacaea (Family)	Unknown Genus and Species	0.235	5.47E-12	0.765	-0.108	0.162	0.945
4408801	Ruminococcacaea (Family)	Oscillospira (unknown species)	0.235	1.32E-08	0.765	0.415	-0.117	0.702
Collapsed	Mollicutes (Class)	RF39 (Order)	0.234	0.0222	0.744	0.292	-0.054	0.762
Collapsed	RF39 (Order)	Unknown Family	0.234	0.0222	0.744	0.292	-0.054	0.762
Collapsed	RF39 (Order)	Unknown Family and Genus	0.234	0.0222	0.744	0.292	-0.054	0.762
Collapsed	Lachnospiraceae (Family)	Unknown Genus	0.233	0	0.767	0.209	0.026	0.766
2724175	Lachnospiraceae (Family)	Unknown Genus and Species	0.233	8.39E-15	0.767	0.067	-0.026	0.960
178845	Ruminococcacaea (Family)	Unknown Genus and Species	0.231	0.103	0.666	0.122	0.174	0.704
181074	Erysipelotrichaceae (Family)	cc_115 (unknown species)	0.231	0	0.769	0.432	-0.168	0.736
305318	Lachnospiraceae (Family)	Unknown Genus and Species	0.231	2.77E-13	0.769	0.204	-0.142	0.938
182643	Clostridiaceae (Family)	Unknown Genus and Species	0.230	2.70E-11	0.770	0.278	-0.084	0.807
2835813	Ruminococcacaea (Family)	Unknown Genus and Species	0.230	3.18E-14	0.770	0.089	-0.036	0.947
Collapsed	Euryarchaeota (Phylum)	Methanobacteria (Class)	0.230	0.107	0.663	0.279	0.047	0.674
Collapsed	Methanobacteria (Class)	Methanobacteriales (Order)	0.230	0.107	0.663	0.279	0.047	0.674
Collapsed	Methanobacteriales (Order)	Methanobacteriaceae (Family)	0.230	0.107	0.663	0.279	0.047	0.674
4428676	Lachnospiraceae (Family)	Coprococcus (unknown species)	0.229	1.01E-11	0.771	0.416	-0.134	0.719
194597	Clostridiales (Order)	Unknown Family, Genus and Species	0.229	2.56E-13	0.771	0.009	0.086	0.904
185420	Bacteroidaceae (Family)	Bacteroides (unknown species)	0.229	6.51E-16	0.771	-0.229	0.286	0.943
4405104	Lachnospiraceae (Family)	Coprococcus (unknown species)	0.229	4.17E-12	0.771	0.344	-0.129	0.784
174818	Ruminococcacaea (Family)	Unknown Genus and Species	0.228	3.23E-14	0.772	0.424	-0.162	0.738
188676	Ruminococcacaea (Family)	Unknown Genus and Species	0.227	0	0.773	0.104	-0.053	0.949
197624	Ruminococcacaea (Family)	Unknown Genus and Species	0.227	3.31E-15	0.773	0.348	-0.170	0.822
189092	Ruminococcacaea (Family)	Faecalibacterium prausnitzii	0.226	8.25E-14	0.774	0.200	-0.123	0.923
178018	Lachnospiraceae (Family)	Unknown Genus and Species	0.225	1.86E-15	0.775	0.368	-0.095	0.727
195937	Lachnospiraceae (Family)	Blautia (unknown species)	0.225	1.01E-13	0.775	0.585	-0.387	0.802

681370	Bifidobacteriaceae (Family)	Bifidobacterium pseudolongum	0.223	4.77E-14	0.777	0.699	-0.385	0.686
316732	Lachnospiraceae (Family)	Lachnospira (unknown species)	0.222	6.73E-12	0.778	0.319	-0.078	0.758
192720	Ruminococcacaea (Family)	Unknown Genus and Species	0.220	0	0.780	0.081	0.053	0.866
337511	Clostridiaceae (Family)	Unknown Genus and Species	0.219	0	0.781	0.424	-0.124	0.700
4438983	Methanobacteriaceae (Family)	Methanobrevibacter (unknown species)	0.218	0.105	0.677	0.272	0.040	0.688
174019	Lachnospiraceae (Family)	Coprococcus (unknown species)	0.218	0.0217	0.760	0.226	-0.006	0.780
Collapsed	Clostridiales (Order)	Lachnospiraceae (Family)	0.217	0.0198	0.763	0.232	-0.006	0.774
4419459	Clostridiales (Order)	Unknown Family, Genus and Species	0.216	0.0163	0.767	0.269	-0.016	0.747
327218	Ruminococcacaea (Family)	Unknown Genus and Species	0.216	2.36E-12	0.784	0.247	0.004	0.749
308544	Ruminococcacaea (Family)	Unknown Genus and Species	0.215	0.0482	0.737	0.253	-0.129	0.876
4453501	Veillonellaceae (Family)	Veillonella dispar	0.214	0.0304	0.756	0.246	0.016	0.738
193709	Ruminococcacaea (Family)	Unknown Genus and Species	0.212	9.87E-16	0.788	-0.297	0.283	1.015
186687	Lachnospiraceae (Family)	Unknown Genus and Species	0.209	1.23E-17	0.791	-0.330	0.264	1.066
2256425	Christensenellaceae (Family)	Unknown Genus and Species	0.208	2.72E-12	0.792	0.085	0.036	0.878
369014	Ruminococcacaea (Family)	Unknown Genus and Species	0.208	1.09E-14	0.792	0.357	-0.142	0.785
4374302	Lachnospiraceae (Family)	Dorea (unknown species)	0.208	1.22E-13	0.792	0.512	-0.282	0.770
4398588	Ruminococcacaea (Family)	Unknown Genus and Species	0.208	5.43E-17	0.792	0.427	-0.184	0.757
Collapsed	Bacilli (Class)	Turicibacterales (Order)	0.208	0.109	0.683	0.274	0.032	0.694
Collapsed	Turicibacterales (Order)	Turicibacteraceae (Family)	0.208	0.109	0.683	0.274	0.032	0.694
Collapsed	Turicibacteraceae (Family)	Turicibacter (Genus)	0.208	0.109	0.683	0.274	0.032	0.694
1602805	Lachnospiraceae (Family)	Unknown Genus and Species	0.208	1.47E-12	0.792	0.194	0.030	0.776
368490	Turicibacteraceae (Family)	Turibacter (unknown species)	0.207	0.0973	0.695	0.268	0.025	0.707
196513	Ruminococcacaea (Family)	Unknown Genus and Species	0.207	6.12E-14	0.793	-0.063	0.095	0.968
179744	Clostridiales (Order)	Unknown Family, Genus and Species	0.207	0.0471	0.746	0.233	-0.002	0.770
196054	Lachnospiraceae (Family)	Unknown Genus and Species	0.205	0.0212	0.774	0.330	-0.088	0.759
4465124	Clostridiaceae (Family)	Clostridium (unknown species)	0.204	7.63E-14	0.796	0.691	-0.414	0.722
332185	Ruminococcacaea (Family)	Unknown Genus and Species	0.204	8.53E-14	0.796	0.291	-0.078	0.787
185763	Ruminococcacaea (Family)	Faecalibacterium prausnitzii	0.203	3.48E-12	0.797	0.260	-0.122	0.862
Collapsed	Clostridiales (Order)	Ruminococcaceae (Family)	0.202	1.20E-11	0.798	0.262	-0.054	0.792
Collapsed	Methanobacteriaceae (Family)	Methanobrevibacter (Genus)	0.201	0.121	0.678	0.250	0.060	0.690
187180	Ruminococcacaea (Family)	Unknown Genus and Species	0.200	0	0.800	0.045	-0.014	0.968
217109	Christensenellaceae (Family)	Unknown Genus and Species	0.199	7.21E-13	0.801	0.096	-0.035	0.938

195651	Ruminococcacaea (Family)	Unknown Genus and Species	0.198	7.04E-14	0.802	0.118	-0.100	0.981
183439	Ruminococcacaea (Family)	Unknown Genus and Species	0.198	4.99E-10	0.802	-0.208	0.157	1.051
Collapsed	Gammaproteobacteria (Class)	Enterobacteriales (Order)	0.196	0.0684	0.736	0.276	-0.006	0.731
Collapsed	Enterobacteriales (Order)	Enterobacteriaceae (Family)	0.196	0.0684	0.736	0.276	-0.006	0.731
182188	Clostridiales (Order)	Unknown Family, Genus and Species	0.195	3.99E-15	0.805	0.017	0.089	0.894
363029	Lachnospiraceae (Family)	Blautia (unknown species)	0.194	1.13E-10	0.806	0.240	-0.045	0.804
192383	Ruminococcacaea (Family)	Ruminococcus (unknown species)	0.193	8.31E-26	0.807	0.097	-0.008	0.911
299302	Lachnospiraceae (Family)	Blautia (unknown species)	0.193	1.40E-15	0.807	0.140	-0.003	0.863
182577	Ruminococcacaea (Family)	Unknown Genus and Species	0.192	0	0.808	0.455	-0.244	0.789
199344	Lachnospiraceae (Family)	Unknown Genus and Species	0.192	0	0.808	0.291	-0.118	0.828
4332082	Lachnospiraceae (Family)	Roseburia (unknown species)	0.192	4.57E-13	0.808	0.220	0.010	0.770
4431545	Ruminococcacaea (Family)	Unknown Genus and Species	0.191	0.0544	0.754	0.179	0.055	0.766
189924	Ruminococcacaea (Family)	Unknown Genus and Species	0.191	1.81E-16	0.809	0.062	0.078	0.860
3973322	Clostridiales (Order)	Unknown Family, Genus and Species	0.190	7.40E-11	0.810	0.498	-0.250	0.751
151870	Erysipelotrichaceae (Family)	Coprobacillus (unknown species)	0.190	3.02E-13	0.810	0.323	-0.100	0.778
302160	Ruminococcacaea (Family)	Unknown Genus and Species	0.189	1.02E-11	0.811	0.261	-0.015	0.754
307113	Lachnospiraceae (Family)	Blautia (unknown species)	0.187	0.0139	0.799	0.068	-0.029	0.961
198626	Lachnospiraceae (Family)	Unknown Genus and Species	0.187	8.96E-14	0.813	0.295	-0.133	0.838
193129	Lachnospiraceae (Family)	Unknown Genus and Species	0.187	0.0104	0.803	0.005	0.015	0.980
4402903	Clostridiales (Order)	Unknown Family, Genus and Species	0.186	1.78E-13	0.814	0.441	-0.220	0.779
178183	Clostridiaceae (Family)	Unknown Genus and Species	0.186	5.24E-26	0.814	0.167	0.047	0.786
4428714	RF39 (Order)	Unknown Family, Genus and Species	0.186	0.0172	0.797	0.121	0.071	0.808
198127	Lachnospiraceae (Family)	Unknown Genus and Species	0.185	1.80E-11	0.815	0.021	0.102	0.877
538322	Ruminococcacaea (Family)	Unknown Genus and Species	0.184	0.00521	0.810	0.257	-0.049	0.792
3265161	Lachnospiraceae (Family)	Unknown Genus and Species	0.184	1.97E-14	0.816	0.207	-0.020	0.813
4094259	Ruminococcacaea (Family)	Unknown Genus and Species	0.183	3.65E-10	0.817	0.350	-0.129	0.778
186732	Ruminococcacaea (Family)	Unknown Genus and Species	0.182	1.15E-11	0.818	0.134	-0.065	0.930
1614788	Lachnospiraceae (Family)	Coprococcus catus	0.182	8.04E-13	0.818	0.091	0.043	0.865
4465907	Lachnospiraceae (Family)	Blautia (unknown species)	0.181	7.39E-12	0.819	0.299	-0.090	0.791
192127	Lachnospiraceae (Family)	Unknown Genus and Species	0.181	3.37E-14	0.819	0.026	0.066	0.907
Collapsed	Proteobacteria (Phylum)	Gammaproteobacteria (Class)	0.180	0.151	0.669	0.168	0.155	0.677
197890	Ruminococcacaea (Family)	Oscillospira (unknown species)	0.179	0	0.821	0.110	0.031	0.860

Collapsed	Ruminococcaceae (Family)	Faecalibacterium (Genus)	0.179	3.80E-13	0.821	0.317	-0.135	0.818
190649	Ruminococcacaea (Family)	Unknown Genus and Species	0.179	2.36E-12	0.821	0.298	-0.109	0.811
187504	Ruminococcacaea (Family)	Unknown Genus and Species	0.177	4.33E-14	0.823	0.410	-0.192	0.782
2506486	Ruminococcacaea (Family)	Unknown Genus and Species	0.177	0.0518	0.771	0.120	-0.063	0.944
352347	Ruminococcacaea (Family)	Unknown Genus and Species	0.176	2.04E-12	0.824	0.397	-0.185	0.788
3211875	Bacteroidaceae (Family)	Bacteroides (unknown species)	0.176	1.37E-12	0.824	0.299	-0.106	0.807
318970	Lachnospiraceae (Family)	Blautia (unknown species)	0.175	0.00392	0.821	0.119	0.060	0.822
179200	Ruminococcacaea (Family)	Unknown Genus and Species	0.174	1.44E-14	0.826	0.425	-0.199	0.773
197970	Ruminococcacaea (Family)	Unknown Genus and Species	0.174	5.75E-13	0.826	0.148	-0.105	0.958
195436	Ruminococcacaea (Family)	Unknown Genus and Species	0.173	2.23E-13	0.827	0.045	0.021	0.934
300620	Ruminococcacaea (Family)	Unknown Genus and Species	0.173	0.00456	0.823	0.643	-0.319	0.677
176113	Clostridiales (Order)	Unknown Family, Genus and Species	0.173	4.11E-14	0.827	0.239	-0.040	0.801
2688035	Lachnospiraceae (Family)	Unknown Genus and Species	0.172	2.35E-10	0.828	-0.096	0.131	0.966
Collapsed	Proteobacteria (Phylum)	Deltaproteobacteriaceae (Family)	0.172	5.23E-10	0.828	0.221	-0.050	0.828
Collapsed	Deltaproteobacteriaceae (Family)	Desulfovibrionales (Order)	0.172	5.18E-12	0.828	0.221	-0.050	0.828
Collapsed	Desulfovibrionales (Order)	Desulfovibrionaceae (Family)	0.172	5.18E-12	0.828	0.221	-0.050	0.828
180473	Ruminococcacaea (Family)	Unknown Genus and Species	0.171	9.08E-15	0.829	0.156	0.017	0.827
4449427	Lachnospiraceae (Family)	Dorea (unknown species)	0.171	5.76E-12	0.829	0.346	-0.125	0.779
313524	Ruminococcacaea (Family)	Unknown Genus and Species	0.170	3.48E-13	0.830	0.062	0.041	0.898
174840	Ruminococcacaea (Family)	Unknown Genus and Species	0.170	0.0143	0.816	0.099	0.063	0.838
174911	Lachnospiraceae (Family)	Unknown Genus and Species	0.170	0.0713	0.759	0.189	0.055	0.756
190162	Lachnospiraceae (Family)	Blautia (unknown species)	0.169	7.35E-15	0.831	-0.281	0.225	1.056
326482	Prevotellaceae (Family)	Prevotella copri	0.168	0.0111	0.820	0.188	0.007	0.805
302049	Lachnospiraceae (Family)	Blautia (unknown species)	0.167	0.0433	0.789	0.027	0.119	0.855
4372528	Clostridiales (Order)	Unknown Family, Genus and Species	0.167	0.124	0.708	0.332	0.000	0.668
328544	Ruminococcacaea (Family)	Unknown Genus and Species	0.167	8.59E-14	0.833	0.526	-0.279	0.753
Collapsed	Proteobacteria (Phylum)	Alphaproteobacteria (Class)	0.167	2.65E-12	0.833	0.285	-0.122	0.837
185814	Ruminococcacaea (Family)	Unknown Genus and Species	0.166	4.79E-14	0.834	0.276	-0.149	0.874
176300	Lachnospiraceae (Family)	Coprococcus (unknown species)	0.166	6.49E-14	0.834	0.029	0.084	0.886
179486	Ruminococcacaea (Family)	Unknown Genus and Species	0.166	1.57E-31	0.834	0.280	-0.074	0.793
182538	Lachnospiraceae (Family)	Coprococcus (unknown species)	0.166	6.02E-13	0.834	0.175	-0.085	0.911
184114	Ruminococcacaea (Family)	Unknown Genus and Species	0.166	3.80E-12	0.834	0.083	0.003	0.914

189588	Lachnospiraceae (Family)	Blautia (unknown species)	0.165	0	0.835	0.192	-0.050	0.857
183604	Lachnospiraceae (Family)	Blautia (unknown species)	0.165	0.0184	0.817	0.230	-0.056	0.825
176604	Lachnospiraceae (Family)	Unknown Genus and Species	0.163	0	0.837	0.068	0.062	0.870
336559	Bacteroidaceae (Family)	Bacteroides (unknown species)	0.162	3.26E-13	0.838	0.388	-0.151	0.763
Collapsed	Clostridiaceae (Family)	Sarcina (Genus)	0.162	0	0.838	0.227	-0.087	0.860
176104	Ruminococcacaea (Family)	Oscillospira (unknown species)	0.162	1.37E-14	0.838	0.427	-0.188	0.761
352747	Lachnospiraceae (Family)	Coprococcus (unknown species)	0.161	7.51E-14	0.839	0.227	-0.058	0.830
191214	Lachnospiraceae (Family)	Unknown Genus and Species	0.160	8.28E-12	0.840	0.293	-0.121	0.828
184525	Lachnospiraceae (Family)	Coprococcus (unknown species)	0.160	1.73E-11	0.840	0.347	-0.163	0.816
Collapsed	Peptococcaceae (Family)	Unknown Genus	0.160	6.21E-13	0.840	0.261	-0.097	0.836
295258	Ruminococcacaea (Family)	Unknown Genus and Species	0.160	3.18E-13	0.840	0.142	0.063	0.794
3422630	Ruminococcacaea (Family)	Unknown Genus and Species	0.159	2.03E-25	0.841	0.329	-0.138	0.809

Table A2: Bonferonni-significant associations between the human microbiome and adiposity.

OTU	Parent Taxonomy (Level)	Taxonomy (Level)	Pheno	P Value	Std Err	Estimate
207487	Lachnospiraceae (Family)	Lachnospira (Unknown Species)	VFM	1.82E-09	0.0053	-0.0324
199344	Lachnospiraceae (Family)	Unknown Genus and Species	GTM	3.46E-08	0.0414	0.2352
199344	Lachnospiraceae (Family)	Unknown Genus and Species	GLM	4.46E-08	0.0173	0.0971
4349261	Lachnospiraceae (Family)	Unknown Genus and Species	pTF	9.09E-08	0.0342	-0.1844
207487	Lachnospiraceae (Family)	Lachnospira (Unknown Species)	AFM	2.08E-07	0.0173	-0.0907
Collapsed	Tenericutes (Phylum)		pTF	2.24E-07	0.0088	-0.0463
289734	Lachnospiraceae (Family)	Unknown Genus and Species	VFM	2.30E-07	0.0014	0.0071
176269	Lachnospiraceae (Family)	Lachnospira (Unknown Species)	VFM	2.34E-07	0.0014	-0.0075
199344	Lachnospiraceae (Family)	Unknown Genus and Species	BMI	2.42E-07	0.0560	0.2957
193769	Lachnospiraceae (Family)	Ruminococcus gnavus	ATM	2.45E-07	0.0167	0.0879
193769	Lachnospiraceae (Family)	Ruminococcus gnavus	GLM	2.85E-07	0.0095	0.0497
193769	Lachnospiraceae (Family)	Ruminococcus gnavus	GTM	2.88E-07	0.0226	0.1197
Collapsed	Tenericutes (Phylum)	Mollicutes (Class)	pTF	5.17E-07	0.0098	-0.0501
322835	Lachnospiraceae (Family)	Lachnospira (Unknown Species)	VFM	6.00E-07	0.0054	-0.0272
199344	Lachnospiraceae (Family)	Unknown Genus and Species	ATM	6.19E-07	0.0308	0.1577
Collapsed	Tenericutes (Phylum)		VFM	6.76E-07	0.0011	-0.0053
Collapsed	Mollicutes (Class)	RF39 (Order)	pTF	7.77E-07	0.0105	-0.0527
Collapsed	RF39 (Order)	Unknown Family	pTF	7.77E-07	0.0105	-0.0527
Collapsed	RF39 (Order)	Unknown Family & Genus	pTF	7.77E-07	0.0105	-0.0527
Collapsed	Tenericutes (Phylum)	RF3 (Class)	VFM	8.13E-07	0.0035	-0.0177
Collapsed	RF3 (Order)	ML615J.28 (Order)	VFM	8.13E-07	0.0035	-0.0177
Collapsed	ML615J.28	Unknown Family	VFM	8.13E-07	0.0035	-0.0177
Collapsed	ML615J.28	Unknown Family & Genus	VFM	8.13E-07	0.0035	-0.0177

3195723	Ruminococcaceae (Family)	Oscillospira (Unknown Species)	BMI	8.54E-07	0.0372	-0.1846
Collapsed	Clostridiales (Order)	Dehalobacteriaceae (Family)	VFM	8.88E-07	0.0032	-0.0160
Collapsed	Dehalobacteriaceae (Family)	Dehalobacterium (Genus)	VFM	9.78E-07	0.0033	-0.0163
193769	Lachnospiraceae (Family)	Ruminococcus gnavus	AFM	9.95E-07	0.0119	0.0592
289734	Lachnospiraceae (Family)	Unknown Genus and Species	A:G	1.11E-06	0.0114	0.0558
194488	Ruminococcaceae (Family)	Unknown Genus and Species	VFM	1.21E-06	0.0057	-0.0281
316732	Lachnospiraceae (Family)	Lachnospira (Unknown Species)	VFM	1.22E-06	0.0014	-0.0068
176269	Lachnospiraceae (Family)	Lachnospira (Unknown Species)	pTF	1.35E-06	0.0122	-0.0596
4465907	Lachnospiraceae (Family)	Blautia (Unknown Species)	VFM	1.46E-06	0.0055	0.0265
180042	Lachnospiraceae (Family)	Unknown Genus and Species	VFM	1.50E-06	0.0057	-0.0276
Collapsed	Mollicutes (Class)	RF39 (Order)	VFM	2.06E-06	0.0013	-0.0061
Collapsed	RF39 (Order)	Unknown Family	VFM	2.06E-06	0.0013	-0.0061
Collapsed	RF39 (Order)	Unknown Family & Genus	VFM	2.06E-06	0.0013	-0.0061
193769	Lachnospiraceae (Family)	Ruminococcus gnavus	ALM	2.11E-06	0.0061	0.0292
199344	Lachnospiraceae (Family)	Unknown Genus and Species	AFM	2.21E-06	0.0219	0.1064
207487	Lachnospiraceae (Family)	Lachnospira (Unknown Species)	SAT	2.24E-06	0.0461	-0.2205
197481	Lachnospiraceae (Family)	Blautia (Unknown Species)	BMI	2.30E-06	0.0738	0.3529
174911	Lachnospiraceae (Family)	Unknown Genus and Species	pTF	2.41E-06	0.0336	-0.1598
Collapsed	Tenericutes (Phylum)	RF3 (Class)	pTF	2.59E-06	0.0299	-0.1413
Collapsed	RF3 (Order)	ML615J.28 (Order)	pTF	2.59E-06	0.0299	-0.1413
Collapsed	ML615J.28	Unknown Family	pTF	2.59E-06	0.0299	-0.1413
Collapsed	ML615J.28	Unknown Family & Genus	pTF	2.59E-06	0.0299	-0.1413
Collapsed	Clostridiales (Order)	Unknown Family	VFM	2.84E-06	0.0009	-0.0042
Collapsed	Clostridiales (Order)	Unknown Family & Genus	VFM	2.84E-06	0.0009	-0.0042
207487	Lachnospiraceae (Family)	Lachnospira (Unknown Species)	pTF	2.94E-06	0.0455	-0.2146
287790	Ruminococcaceae (Family)	Unknown Genus and Species	GFM	3.03E-06	0.0181	-0.0849

207487	Lachnospiraceae (Family)	Lachnospira (Unknown Species)	ATM	3.25E-06	0.0244	-0.1147
322835	Lachnospiraceae (Family)	Lachnospira (Unknown Species)	pTF	3.31E-06	0.0461	-0.2160
194488	Ruminococcaceae (Family)	Unknown Genus and Species	pTF	3.43E-06	0.0486	-0.2271
287790	Ruminococcaceae (Family)	Unknown Genus and Species	SAT	3.46E-06	0.0365	-0.1704
Collapsed	Christensenellaceae (Family)	Unknown Genus	VFM	3.51E-06	0.0015	-0.0071
3014082	Lachnospiraceae (Family)	Blautia (Unknown Species)	AFM	3.59E-06	0.0093	0.0438
292735	Lachnospiraceae (Family)	Blautia (Unknown Species)	VFM	3.69E-06	0.0029	0.0135
194597	Clostridiales (Order)	Unknown Family, Genus & Species	VFM	3.79E-06	0.0031	-0.0144
199344	Lachnospiraceae (Family)	Unknown Genus and Species	ALM	3.99E-06	0.0111	0.0523
199344	Lachnospiraceae (Family)	Unknown Genus and Species	GFM	4.06E-06	0.0287	0.1353
332929	Ruminococcaceae (Family)	Unknown Genus and Species	AFM	4.13E-06	0.0164	0.0768
176269	Lachnospiraceae (Family)	Lachnospira (Unknown Species)	AFM	4.37E-06	0.0047	-0.0216
176269	Lachnospiraceae (Family)	Lachnospira (Unknown Species)	A:G	4.43E-06	0.0120	-0.0556
146564	Clostridiales (Order)	Unknown Family, Genus & Species	VFM	4.43E-06	0.0027	-0.0126
4349261	Lachnospiraceae (Family)	Unknown Genus and Species	VFM	4.50E-06	0.0041	-0.0191
289734	Lachnospiraceae (Family)	Unknown Genus and Species	pTF	4.54E-06	0.0114	0.0529
332929	Ruminococcaceae (Family)	Unknown Genus and Species	ATM	4.61E-06	0.0231	0.1077

Table A3: Suggestive associations between 97 known obesity loci Locke et al. [2015] and adiposity in an expanded 16S dataset.

CHR	SNP	Position	Locke Gene	AF	P Value	Beta	SE	OTU	Parent Taxonomy (Level)	Taxonomy (Level)
3	rs74331972	61219398	FHIT	0.113	2.49E-06	3.29E-01	6.95E-02	181702	Clostridiales (Order)	Unknown Family, Genus and Species
6	rs1433723	40350399	TDRG1	0.302	4.32E-06	2.23E-01	4.83E-02	OR25576	Clostridiales (Order)	Unknown Family, Genus and Species
6	rs1433722	40350655	TDRG1	0.302	4.89E-06	2.21E-01	4.82E-02	OR25576	Clostridiales (Order)	Unknown Family, Genus and Species
1	rs2480677	50577654	ELAV	0.123	4.95E-06	2.99E-01	6.51E-02	194733	Lachnospiraceae (Family)	Blautia (Unknown Species)
1	rs7517535	50584682	ELAV	0.122	6.31E-06	2.95E-01	6.50E-02	194733	Lachnospiraceae (Family)	Blautia (Unknown Species)
6	rs4145198	40351898	TDRG1	0.304	6.77E-06	2.16E-01	4.77E-02	OR25576	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs1530672	81786460	GBE1	0.473	8.01E-06	1.94E-01	4.32E-02	538322	Ruminococcaceae (Family)	Oscillospira (Unknown Species)
3	rs6799513	61226260	FHIT	0.125	8.42E-06	3.00E-01	6.70E-02	181702	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs6799836	61226759	FHIT	0.125	8.42E-06	3.00E-01	6.70E-02	181702	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs6763941	61226766	FHIT	0.125	8.42E-06	3.00E-01	6.70E-02	181702	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs7640496	61216475	FHIT	0.126	8.67E-06	2.99E-01	6.68E-02	181702	Clostridiales (Order)	Unknown Family, Genus and Species
6	rs4714344	40354999	TDRG1	0.301	1.06E-05	2.11E-01	4.77E-02	OR25576	Clostridiales (Order)	Unknown Family, Genus and Species
6	rs17681421	40357094	TDRG1	0.301	1.06E-05	2.11E-01	4.77E-02	OR25576	Clostridiales (Order)	Unknown Family, Genus and Species
6	rs11758429	40363345	TDRG1	0.301	1.06E-05	2.11E-01	4.77E-02	OR25576	Clostridiales (Order)	Unknown Family, Genus and Species
19	rs398426	34321258	KCTD15	0.277	1.20E-05	-2.06E-01	4.69E-02	289734	Lachnospiraceae (Family)	Unknown Genus and Species
6	rs17619780	40364325	TDRG1	0.301	1.26E-05	2.10E-01	4.77E-02	OR25576	Clostridiales (Order)	Unknown Family, Genus and Species
6	rs2504847	40371199	TDRG1	0.338	1.28E-05	2.04E-01	4.65E-02	OR25576	Clostridiales (Order)	Unknown Family, Genus and Species
13	rs7319716	79559673	MIR548A2	0.494	1.32E-05	-1.94E-01	4.44E-02	196905	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs3772889	81792592	GBE1	0.451	1.37E-05	1.89E-01	4.32E-02	538322	Ruminococcaceae (Family)	Oscillospira (Unknown Species)
3	rs2594553	81793085	GBE1	0.451	1.37E-05	1.89E-01	4.32E-02	538322	Ruminococcaceae (Family)	Oscillospira (Unknown Species)
3	rs2680246	81803293	GBE1	0.451	1.37E-05	1.89E-01	4.32E-02	538322	Ruminococcaceae (Family)	Oscillospira (Unknown Species)
3	rs2680245	81805517	GBE1	0.451	1.37E-05	1.89E-01	4.32E-02	538322	Ruminococcaceae (Family)	Oscillospira (Unknown Species)
3	rs2594558	81811845	GBE1	0.451	1.37E-05	1.89E-01	4.32E-02	538322	Ruminococcaceae (Family)	Oscillospira (Unknown Species)
3	rs2680277	81812406	GBE1	0.451	1.37E-05	1.89E-01	4.32E-02	538322	Ruminococcaceae (Family)	Oscillospira (Unknown Species)
3	rs9863009	81815218	GBE1	0.451	1.37E-05	1.89E-01	4.32E-02	538322	Ruminococcaceae (Family)	Oscillospira (Unknown Species)
3	rs13322666	61220224	FHIT	0.124	1.41E-05	2.93E-01	6.71E-02	181702	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs9829886	61222555	FHIT	0.124	1.41E-05	2.93E-01	6.71E-02	181702	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs9858841	61229611	FHIT	0.124	1.49E-05	2.92E-01	6.71E-02	181702	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs7640756	61216755	FHIT	0.125	1.49E-05	2.92E-01	6.70E-02	181702	Clostridiales (Order)	Unknown Family, Genus and Species
6	rs731575	40369195	TDRG1	0.3	1.54E-05	2.08E-01	4.79E-02	OR25576	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs2594551	81790290	GBE1	0.451	1.72E-05	1.87E-01	4.33E-02	538322	Ruminococcaceae (Family)	Oscillospira (Unknown Species)
13	rs4421883	79561962	MIR548A2	0.497	1.89E-05	-1.92E-01	4.48E-02	196905	Ruminococcaceae (Family)	Unknown Genus and Species

3	rs1530672	81786460	GBE1	0.473	2.21E-05	1.89E-01	4.43E-02	4450214	Ruminococcaceae (Family)	Unknown Genus and Species
13	rs944512	79565016	MIR548A2	0.497	2.32E-05	-1.91E-01	4.49E-02	196905	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs3772889	81792592	GBE1	0.451	2.44E-05	1.88E-01	4.44E-02	4450214	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs2594553	81793085	GBE1	0.451	2.44E-05	1.88E-01	4.44E-02	4450214	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs2680246	81803293	GBE1	0.451	2.44E-05	1.88E-01	4.44E-02	4450214	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs2680245	81805517	GBE1	0.451	2.44E-05	1.88E-01	4.44E-02	4450214	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs2594558	81811845	GBE1	0.451	2.44E-05	1.88E-01	4.44E-02	4450214	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs2680277	81812406	GBE1	0.451	2.44E-05	1.88E-01	4.44E-02	4450214	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs9863009	81815218	GBE1	0.451	2.44E-05	1.88E-01	4.44E-02	4450214	Ruminococcaceae (Family)	Unknown Genus and Species
13	rs7999857	79558885	MIR548A2	0.495	2.47E-05	-1.89E-01	4.46E-02	196905	Ruminococcaceae (Family)	Unknown Genus and Species
6	rs13207511	108953975	$FOXO3$	0.263	2.54E-05	-2.13E-01	5.04E-02	190864	Clostridiales (Order)	Unknown Family, Genus and Species
16	rs1861869	53790181	FTO	0.49	2.56E-05	-1.87E-01	4.41E-02	581201	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs200399408	81770407	GBE1	0.45	2.66E-05	1.82E-01	4.31E-02	538322	Ruminococcaceae (Family)	Oscillospira (Unknown Species)
3	rs149763534	81770408	GBE1	0.45	2.66E-05	1.82E-01	4.31E-02	538322	Ruminococcaceae (Family)	Oscillospira (Unknown Species)
3	rs1155277	81771587	GBE1	0.45	2.66E-05	1.82E-01	4.31E-02	538322	Ruminococcaceae (Family)	Oscillospira (Unknown Species)
3	rs3107240	81774668	GBE1	0.45	2.66E-05	1.82E-01	4.31E-02	538322	Ruminococcaceae (Family)	Oscillospira (Unknown Species)
3	rs2594547	81777680	GBE1	0.45	2.66E-05	1.82E-01	4.31E-02	538322	Ruminococcaceae (Family)	Oscillospira (Unknown Species)
9	rs3861877	129446591	LMX1B	0.291	2.76E-05	2.16E-01	5.12E-02	199421	Lachnospiraceae (Family)	Lachnospira (Unknown Species)
13	rs9601149	79566442	MIR548A2	0.497	2.84E-05	-1.89E-01	4.49E-02	196905	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs3836324	81775388	GBE1	0.45	2.85E-05	1.82E-01	4.32E-02	538322	Ruminococcaceae (Family)	Oscillospira (Unknown Species)
3	rs2680257	81781813	GBE1	0.45	2.86E-05	1.82E-01	4.32E-02	538322	Ruminococcaceae (Family)	Oscillospira (Unknown Species)
3	rs2594549	81785355	GBE1	0.45	2.86E-05	1.82E-01	4.32E-02	538322	Ruminococcaceae (Family)	Oscillospira (Unknown Species)
3	rs60530974	81786888	GBE1	0.45	2.86E-05	1.82E-01	4.32E-02	538322	Ruminococcaceae (Family)	Oscillospira (Unknown Species)
17	rs148116771	5267160	RABEP1	0.121	2.87E-05	3.04E-01	7.24E-02	182196	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs3772889	81792592	GBE1	0.451	2.93E-05	1.94E-01	4.62E-02	553080	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs2594553	81793085	GBE1	0.451	2.93E-05	1.94E-01	4.62E-02	553080	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs2680246	81803293	GBE1	0.451	2.93E-05	1.94E-01	4.62E-02	553080	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs2680245	81805517	GBE1	0.451	2.93E-05	1.94E-01	4.62E-02	553080	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs2594558	81811845	GBE1	0.451	2.93E-05	1.94E-01	4.62E-02	553080	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs2680277	81812406	GBE1	0.451	2.93E-05	1.94E-01	4.62E-02	553080	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs9863009	81815218	GBE1	0.451	2.93E-05	1.94E-01	4.62E-02	553080	Clostridiales (Order)	Unknown Family, Genus and Species
6	rs4294001	50850065	TFAP2B	0.118	3.03E-05	-2.90E-01	6.92E-02	313524	Ruminococcaceae (Family)	Unknown Genus and Species
8	rs1462436	76804524	HNF4G	0.12	3.10E-05	-2.69E-01	6.43E-02	4321810	Lachnospiraceae (Family)	Blautia (Unknown Species)
3	rs6771266	81787052	GBE1	0.452	3.12E-05	1.81E-01	4.33E-02	538322	Ruminococcaceae (Family)	Oscillospira (Unknown Species)
3	rs11712466	81770363	GBE1	0.463	3.19E-05	1.87E-01	4.48E-02	4450214	Ruminococcaceae (Family)	Unknown Genus and Species
6	rs7738924	50850600	TFAP2B	0.116	3.26E-05	-2.92E-01	7.00E-02	313524	Ruminococcaceae (Family)	Unknown Genus and Species

6	rs9473931	50851411	TFAP2B	0.116	3.26E-05	-2.92E-01	7.00E-02	313524	Ruminococcaceae (Family)	Unknown Genus and Species
6	rs7746581	50853479	TFAP2B	0.116	3.26E-05	-2.92E-01	7.00E-02	313524	Ruminococcaceae (Family)	Unknown Genus and Species
6	rs7753025	50858352	TFAP2B	0.116	3.26E-05	-2.92E-01	7.00E-02	313524	Ruminococcaceae (Family)	Unknown Genus and Species
6	rs71570523	50859119	TFAP2B	0.116	3.26E-05	-2.92E-01	7.00E-02	313524	Ruminococcaceae (Family)	Unknown Genus and Species
6	rs7745565	50860083	TFAP2B	0.116	3.26E-05	-2.92E-01	7.00E-02	313524	Ruminococcaceae (Family)	Unknown Genus and Species
6	rs150926001	50863218	TFAP2B	0.116	3.26E-05	-2.92E-01	7.00E-02	313524	Ruminococcaceae (Family)	Unknown Genus and Species
6	rs6908879	50869856	TFAP2B	0.116	3.26E-05	-2.92E-01	7.00E-02	313524	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs2594551	81790290	GBE1	0.451	3.28E-05	1.86E-01	4.45E-02	4450214	Ruminococcaceae (Family)	Unknown Genus and Species
9	rs10987405	129441797	LMX1B	0.297	3.34E-05	2.13E-01	5.12E-02	199421	Lachnospiraceae (Family)	Lachnospira (Unknown Species)
3	rs1530672	81786460	GBE1	0.473	3.36E-05	1.91E-01	4.59E-02	553080	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs11712466	81770363	GBE1	0.463	3.45E-05	1.82E-01	4.39E-02	538322	Ruminococcaceae (Family)	Oscillospira (Unknown Species)
13	rs9530844	79570300	MIR548A2	0.498	3.57E-05	-1.86E-01	4.49E-02	196905	Ruminococcaceae (Family)	Unknown Genus and Species
2	rs6728479	654472	TMEM18	0.077	3.58E-05	2.85E-01	6.87E-02	OR26796	Lachnospiraceae (Family)	Lachnospira (Unknown Species)
3	rs9858680	61229542	FHIT	0.158	3.64E-05	2.47E-01	5.96E-02	181702	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs11712466	81770363	GBE1	0.463	3.70E-05	1.93E-01	4.65E-02	553080	Clostridiales (Order)	Unknown Family, Genus and Species
6	rs34793301	50859462	TFAP2B	0.116	3.86E-05	-2.90E-01	7.01E-02	313524	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs2594551	81790290	GBE1	0.451	3.91E-05	1.91E-01	4.62E-02	553080	Clostridiales (Order)	Unknown Family, Genus and Species
9	rs7030112	129445436	LMX1B	0.292	3.96E-05	2.11E-01	5.11E-02	199421	Lachnospiraceae (Family)	Lachnospira (Unknown Species)
3	rs1125285	61228797	FHIT	0.158	4.30E-05	2.45E-01	5.96E-02	181702	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs2680264	81779205	GBE1	0.451	4.53E-05	1.77E-01	4.33E-02	538322	Ruminococcaceae (Family)	Oscillospira (Unknown Species)
3	rs2680264	81779205	GBE1	0.451	4.57E-05	1.82E-01	4.45E-02	4450214	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs2680257	81781813	GBE1	0.45	4.63E-05	1.81E-01	4.44E-02	4450214	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs2594549	81785355	GBE1	0.45	4.63E-05	1.81E-01	4.44E-02	4450214	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs60539974	81786888	GBE1	0.45	4.63E-05	1.81E-01	4.44E-02	4450214	Ruminococcaceae (Family)	Unknown Genus and Species
9	rs3780332	129454714	LMX1B	0.296	4.83E-05	2.07E-01	5.08E-02	199421	Lachnospiraceae (Family)	Lachnospira (Unknown Species)
3	rs200399408	81770407	GBE1	0.45	4.92E-05	1.81E-01	4.43E-02	4450214	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs149763534	81770408	GBE1	0.45	4.92E-05	1.81E-01	4.43E-02	4450214	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs1155277	81771587	GBE1	0.45	4.92E-05	1.81E-01	4.43E-02	4450214	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs3107240	81774668	GBE1	0.45	4.92E-05	1.81E-01	4.43E-02	4450214	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs2594547	81777680	GBE1	0.45	4.92E-05	1.81E-01	4.43E-02	4450214	Ruminococcaceae (Family)	Unknown Genus and Species
9	rs35980483	129443852	LMX1B	0.293	5.13E-05	2.07E-01	5.10E-02	199421	Lachnospiraceae (Family)	Lachnospira (Unknown Species)
3	rs3836324	81775388	GBE1	0.45	5.32E-05	1.80E-01	4.44E-02	4450214	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs74331972	61219398	FHIT	0.113	5.38E-05	2.86E-01	7.04E-02	287790	Ruminococcaceae (Family)	Unknown Genus and Species
16	rs1558760	49039077	CBLN1	0.38	5.38E-05	1.82E-01	4.50E-02	182797	Ruminococcaceae (Family)	Unknown Genus and Species
16	rs2908868	49039308	CBLN1	0.38	5.38E-05	1.82E-01	4.50E-02	182797	Ruminococcaceae (Family)	Unknown Genus and Species
13	rs12584042	79563234	MIR548A2	0.484	5.56E-05	-1.81E-01	4.48E-02	196905	Ruminococcaceae (Family)	Unknown Genus and Species

17	rs75754694	5259765	RABEP1	0.122	5.65E-05	2.92E-01	7.21E-02	182196	Clostridiales (Order)	Unknown Family, Genus and Species
6	rs35058096	40346055	TDRG1	0.297	5.68E-05	1.98E-01	4.90E-02	OR25576	Clostridiales (Order)	Unknown Family, Genus and Species
6	rs2304285	40346272	TDRG1	0.297	5.68E-05	1.98E-01	4.90E-02	OR25576	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs2680257	81781813	GBE1	0.45	5.77E-05	1.86E-01	4.61E-02	553080	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs2594549	81785355	GBE1	0.45	5.77E-05	1.86E-01	4.61E-02	553080	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs60530974	81786888	GBE1	0.45	5.77E-05	1.86E-01	4.61E-02	553080	Clostridiales (Order)	Unknown Family, Genus and Species
16	rs1477099	49041694	CBLN1	0.379	6.12E-05	1.81E-01	4.51E-02	182797	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs200399408	81770407	GBE1	0.45	6.28E-05	1.85E-01	4.61E-02	553080	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs149763534	81770408	GBE1	0.45	6.28E-05	1.85E-01	4.61E-02	553080	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs1155277	81771587	GBE1	0.45	6.28E-05	1.85E-01	4.61E-02	553080	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs3107240	81774668	GBE1	0.45	6.28E-05	1.85E-01	4.61E-02	553080	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs2594547	81777680	GBE1	0.45	6.28E-05	1.85E-01	4.61E-02	553080	Clostridiales (Order)	Unknown Family, Genus and Species
16	rs715017	49039976	CBLN1	0.379	6.37E-05	1.81E-01	4.51E-02	182797	Ruminococcaceae (Family)	Unknown Genus and Species
16	rs755284	49044242	CBLN1	0.378	6.42E-05	1.81E-01	4.51E-02	182797	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs3836324	81775388	GBE1	0.45	6.64E-05	1.85E-01	4.61E-02	553080	Clostridiales (Order)	Unknown Family, Genus and Species
6	rs9473921	50828072	TFAP2B	0.117	6.65E-05	-2.79E-01	6.97E-02	313524	Ruminococcaceae (Family)	Unknown Genus and Species
6	rs35546156	50829785	TFAP2B	0.117	6.65E-05	-2.79E-01	6.97E-02	313524	Ruminococcaceae (Family)	Unknown Genus and Species
6	rs9296645	50831194	TFAP2B	0.117	6.65E-05	-2.79E-01	6.97E-02	313524	Ruminococcaceae (Family)	Unknown Genus and Species
8	rs60667200	76808121	HNF4G	0.102	6.85E-05	-2.79E-01	6.98E-02	4321810	Lachnospiraceae (Family)	Blautia (Unknown Species)
13	rs9544947	79563547	MIR548A2	0.484	6.88E-05	-1.79E-01	4.48E-02	196905	Ruminococcaceae (Family)	Unknown Genus and Species
3	rs6771266	81787052	GBE1	0.452	6.96E-05	1.84E-01	4.62E-02	553080	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs3158	85793588	CADM2	0.262	7.00E-05	-1.98E-01	4.95E-02	584107	Clostridiales (Order)	Unknown Family, Genus and Species
17	rs78813154	5267002	RABEP1	0.144	7.27E-05	2.69E-01	6.74E-02	182196	Clostridiales (Order)	Unknown Family, Genus and Species
16	rs149709303	49065242	CBLN1	0.38	7.68E-05	1.82E-01	4.59E-02	182797	Ruminococcaceae (Family)	Unknown Genus and Species
16	rs7184874	53792439	FTO	0.486	7.70E-05	1.76E-01	4.44E-02	581201	Ruminococcaceae (Family)	Unknown Genus and Species
11	rs11042033	8697054	TRIM66	0.169	7.72E-05	2.32E-01	5.84E-02	OR9470	Clostridiales (Order)	Unknown Family, Genus and Species
3	rs2680264	81779205	GBE1	0.451	7.79E-05	1.83E-01	4.62E-02	553080	Clostridiales (Order)	Unknown Family, Genus and Species
8	rs145556455	76787018	HNF4G	0.101	7.96E-05	-2.77E-01	7.00E-02	4321810	Lachnospiraceae (Family)	Blautia (Unknown Species)
2	rs13006797	213405952	ERBB4	0.139	7.99E-05	2.64E-01	6.65E-02	180585	Clostridiales (Order)	Unknown Family, Genus and Species
16	rs2908890	49053964	CBLN1	0.372	8.21E-05	1.82E-01	4.59E-02	182797	Ruminococcaceae (Family)	Unknown Genus and Species
13	rs9601154	79574911	MIR548A2	0.054	8.35E-05	3.69E-01	9.35E-02	321453	Lachnospiraceae (Family)	Blautia (Unknown Species)
3	rs6771266	81787052	GBE1	0.452	8.65E-05	1.75E-01	4.44E-02	4450214	Ruminococcaceae (Family)	Unknown Genus and Species
6	rs34598277	40346864	TDRG1	0.297	8.72E-05	1.93E-01	4.90E-02	OR25576	Clostridiales (Order)	Unknown Family, Genus and Species
13	rs7995741	28039440	MTIF3	0.468	8.80E-05	-1.72E-01	4.36E-02	OR15088	Ruminococcaceae (Family)	Unknown Genus and Species
11	rs34655699	8660117	TRIM66	0.157	9.26E-05	2.38E-01	6.05E-02	OR9470	Clostridiales (Order)	Unknown Family, Genus and Species
16	rs3095670	49048972	CBLN1	0.377	9.30E-05	1.77E-01	4.50E-02	182797	Ruminococcaceae (Family)	Unknown Genus and Species

6	rs36093098	50833931	TFAP2B	0.119	9.83E-05	-2.72E-01	6.97E-02	313524	Ruminococcaceae (Family)	Unknown Genus and Species
6	rs9296644	50828247	TFAP2B	0.118	9.96E-05	-2.70E-01	6.92E-02	313524	Ruminococcaceae (Family)	Unknown Genus and Species

Table A4: 111 Bonferonni-significant associations between 16S gut microbiome and serum/plasma metabolites.

OTU	Parent Taxonomy (Level)	Taxonomy (Level)	Metabolite	P Value	Estimate	SE
Collapsed	Clostridiales (Order)	Christensenellaceae (Family)	palmitate	3.44E-26	0.133	0.012
Collapsed	Christensenellaceae (Family)	Unknown Genus	palmitate	2.14E-25	0.129	0.012
Collapsed	Clostridiales (Order)	Christensenellaceae (Family)	phenylalanine	3.60E-17	0.117	0.014
Collapsed	Christensenellaceae (Family)	Unknown Genus	phenylalanine	9.42E-17	0.114	0.013
4372528	Clostridiales (Order)	Unknown Family, Genus and Species	cholesterol	2.49E-15	0.181	0.022
231952	Christensenellaceae (Family)	Unknown Genus and Species	palmitate	9.04E-15	0.367	0.046
Collapsed	Clostridiales (Order)	Dehalobacteriaceae (Family)	palmitate	1.16E-14	0.211	0.027
4383953	Clostridiaceae (Family)	Unknown Genus and Species	cholesterol	1.26E-14	0.121	0.015
Collapsed	Dehalobacteriaceae (Family)	Dehalobacterium (Genus)	palmitate	3.86E-14	0.213	0.028
289734	Lachnospiraceae (Family)	Unknown Genus and Species	palmitate	5.01E-14	-0.087	0.011
4434334	Clostridiaceae (Family)	Unknown Genus and Species	cholesterol	5.62E-14	0.081	0.011
Collapsed	Peptostreptococcaceae (Family)	Unknown Genus	cholesterol	6.36E-14	0.331	0.043
3195723	Ruminococcaceae (Family)	Oscillospira (Unknown Species)	palmitate	9.31E-14	0.292	0.039
Collapsed	Clostridiales (Order)	Peptostreptococcaceae (Family)	cholesterol	1.41E-13	0.316	0.041
292735	Lachnospiraceae (Family)	Blautia (Unknown Species)	palmitate	2.07E-13	-0.184	0.024
Collapsed	Clostridiaceae (Family)	Unknown Genus	cholesterol	2.13E-13	0.128	0.017
759816	Ruminococcaceae (Family)	Unknown Genus and Species	saccharin	8.45E-13	0.399	0.055
Collapsed	Clostridiaceae (Family)	SMB53 (Genus)	cholesterol	1.83E-12	0.156	0.021
190162	Lachnospiraceae (Family)	Blautia (Unknown Species)	palmitate	2.13E-12	-0.293	0.041
176318	Christensenellaceae (Family)	Unknown Genus and Species	palmitate	4.10E-12	0.098	0.014
Collapsed	Clostridiales (Order)	Lachnospiraceae (Family)	palmitate	1.02E-11	-0.038	0.005
304779	Clostridiaceae (Family)	Clostridium perfringens	cholesterol	1.28E-11	0.371	0.053
308544	Ruminococcaceae (Family)	Unknown Genus and Species	palmitate	1.30E-11	-0.184	0.027
Collapsed	Lachnospiraceae (Family)	Blautia (Genus)	palmitate	1.46E-11	-0.121	0.018
535955	Clostridiaceae (Family)	SMB53 (Unknown Species)	cholesterol	1.82E-11	0.207	0.030
194659	Clostridiaceae (Family)	Unknown Genus and Species	cholesterol	2.03E-11	0.333	0.048

178462	Lachnospiraceae (Family)	Blautia (Unknown Species)	palmitate	6.27E-11	-0.128	0.019
183698	Lachnospiraceae (Family)	Unknown Genus and Species	palmitate	6.61E-11	-0.251	0.038
Collapsed	Lachnospiraceae (Family)	Unknown Genus	palmitate	1.00E-10	-0.059	0.009
4468466	Ruminococcaceae (Family)	Unknown Genus and Species	palmitate	1.24E-10	0.061	0.009
3014082	Lachnospiraceae (Family)	Blautia (Unknown Species)	palmitate	1.29E-10	-0.162	0.025
Collapsed	Clostridiaceae (Family)	Unknown Genus	phenylacetate	1.40E-10	-0.108	0.017
190100	Lachnospiraceae (Family)	Blautia (Unknown Species)	palmitate	3.75E-10	-0.281	0.044
177567	Ruminococcaceae (Family)	Unknown Genus and Species	palmitate	4.61E-10	-0.212	0.034
Collapsed	Bacilli (Class)	Turicibacterales (Order)	cholesterol	5.39E-10	0.152	0.024
Collapsed	Turicibacterales (Order)	Turicibacteriaceae (Family)	cholesterol	5.39E-10	0.152	0.024
Collapsed	Turicibacteriaceae (Family)	Turicibacter (Genus)	cholesterol	5.39E-10	0.152	0.024
4439469	Ruminococcaceae (Family)	Unknown Genus and Species	palmitate	7.05E-10	0.169	0.027
368490	Turicibacteraceae (Family)	Turicibacter (Unknown Species)	cholesterol	7.80E-10	0.153	0.024
181756	Lachnospiraceae (Family)	Blautia (Unknown Species)	palmitate	1.16E-09	-0.322	0.052
183604	Lachnospiraceae (Family)	Blautia (Unknown Species)	palmitate	1.36E-09	-0.150	0.024
3931537	Clostridiaceae (Family)	Clostridium (Unknown Species)	palmitate	1.59E-09	-0.068	0.011
Collapsed	Ruminococcaceae (Family)	Oscillospira (Genus)	palmitate	1.62E-09	0.142	0.023
Collapsed	Clostridiales (Order)	Clostridiaceae (Family)	phenylacetate	2.01E-09	-0.099	0.016
555547	Christensenellaceae (Family)	Unknown Genus and Species	palmitate	3.61E-09	0.207	0.035
4428676	Lachnospiraceae (Family)	Coprococcus (Unknown Species)	hypoxanthine	3.72E-09	-0.159	0.027
176318	Christensenellaceae (Family)	Unknown Genus and Species	phenylalanine	3.82E-09	0.092	0.015
4465907	Lachnospiraceae (Family)	Blautia (Unknown Species)	palmitate	4.70E-09	-0.284	0.048
113003	Ruminococcaceae (Family)	Ruminococcus (Unknown Species)	palmitate	5.38E-09	0.397	0.067
Collapsed	Pasteurellaceae (Family)	Haemophilus (Genus)	cholesterol	5.42E-09	0.151	0.026
289734	Lachnospiraceae (Family)	Unknown Genus and Species	phenylalanine	5.57E-09	-0.075	0.013
Collapsed	Gammaproteobacteria (Class)	Pasteurellales (Order)	cholesterol	5.64E-09	0.149	0.025
Collapsed	Pasteurellales (Order)	Pasteurellaceae (Family)	cholesterol	5.64E-09	0.149	0.025
4477696	Pasteurellaceae (Family)	Haemophilus parainfluenzae	cholesterol	5.81E-09	0.152	0.026
Collapsed	Clostridiales (Order)	Christensenellaceae (Family)	histidine	6.72E-09	0.078	0.013

302049	Lachnospiraceae (Family)	Blautia (Unknown Species)	palmitate	7.38E-09	-0.156	0.027
836693	SHA98 (Order)	Unknown Family, Genus and Species	palmitate	7.64E-09	0.224	0.038
Collapsed	Clostridia (Class)	SHA.98 (Order)	palmitate	7.64E-09	0.224	0.038
Collapsed	SHA.98 (Order)	Unknown Family	palmitate	7.64E-09	0.224	0.038
Collapsed	SHA.98 (Order)	Unknown Family and Genus	palmitate	7.64E-09	0.224	0.038
3265161	Lachnospiraceae (Family)	Unknown Genus and Species	palmitate	8.19E-09	-0.254	0.043
Collapsed	Lachnospiraceae (Family)	Roseburia (Genus)	phenylalanine	8.23E-09	-0.055	0.009
2256425	Christensenellaceae (Family)	Unknown Genus and Species	palmitate	8.97E-09	0.262	0.045
Collapsed	Christensenellaceae (Family)	Unknown Genus	histidine	9.26E-09	0.076	0.013
292735	Lachnospiraceae (Family)	Blautia (Unknown Species)	phenylalanine	1.18E-08	-0.158	0.027
Collapsed	Clostridiales (Order)	Clostridiaceae (Family)	cholesterol	1.31E-08	0.098	0.017
195465	Lachnospiraceae (Family)	Unknown Genus and Species	palmitate	1.38E-08	-0.249	0.043
4370941	Ruminococcaceae (Family)	Unknown Genus and Species	palmitate	1.41E-08	0.241	0.042
176604	Lachnospiraceae (Family)	Unknown Genus and Species	palmitate	1.44E-08	-0.154	0.027
196100	Lachnospiraceae (Family)	Unknown Genus and Species	palmitate	1.54E-08	-0.377	0.066
192983	Lachnospiraceae (Family)	Unknown Genus and Species	palmitate	2.05E-08	-0.184	0.032
332929	Ruminococcaceae (Family)	Unknown Genus and Species	palmitate	2.09E-08	-0.253	0.044
194660	Lachnospiraceae (Family)	Blautia (Unknown Species)	palmitate	2.30E-08	-0.227	0.040
231952	Christensenellaceae (Family)	Unknown Genus and Species	phenylalanine	2.35E-08	0.292	0.052
Collapsed	Bacilli (Class)	Turicibacterales (Order)	phenylacetate	2.41E-08	-0.133	0.023
Collapsed	Turicibacterales (Order)	Turicibacteriaceae (Family)	phenylacetate	2.41E-08	-0.133	0.023
Collapsed	Turicibacteriaceae (Family)	Turicibacter (Genus)	phenylacetate	2.41E-08	-0.133	0.023
191412	Lachnospiraceae (Family)	Unknown Genus and Species	palmitate	2.48E-08	-0.298	0.053
360890	Clostridiales (Order)	Unknown Family, Genus and Species	palmitate	2.52E-08	0.159	0.028
368490	Turicibacteraceae (Family)	Turicibacter (Unknown Species)	phenylacetate	2.79E-08	-0.135	0.024
4336939	Ruminococcaceae (Family)	Unknown Genus and Species	palmitate	2.84E-08	0.162	0.029
3134492	Lachnospiraceae (Family)	Unknown Genus and Species	palmitate	2.85E-08	-0.075	0.013
3134492	Lachnospiraceae (Family)	Unknown Genus and Species	phenylalanine	3.47E-08	-0.082	0.015
1740283	Lachnospiraceae (Family)	Roseburia (Unknown Species)	palmitate	3.89E-08	-0.263	0.047

177515	Lachnospiraceae (Family)	Roseburia (Unknown Species)	phenylalanine	3.94E-08	-0.067	0.012
181174	Ruminococcaceae (Family)	Unknown Genus and Species	palmitate	3.96E-08	0.144	0.026
199702	Ruminococcaceae (Family)	Faecalibacterium prausnitzii	palmitate	4.46E-08	-0.282	0.051
199694	Clostridiaceae (Family)	Clostridium (Unknown Species)	palmitate	4.47E-08	-0.322	0.058
4217963	Lachnospiraceae (Family)	Unknown Genus and Species	palmitate	4.68E-08	-0.312	0.057
295258	Ruminococcaceae (Family)	Unknown Genus and Species	palmitate	5.27E-08	0.123	0.022
197930	Lachnospiraceae (Family)	Unknown Genus and Species	palmitate	5.83E-08	-0.263	0.048
2250985	Lachnospiraceae (Family)	Roseburia (Unknown Species)	palmitate	5.91E-08	-0.199	0.036
1740283	Lachnospiraceae (Family)	Roseburia (Unknown Species)	phenylalanine	6.20E-08	-0.283	0.052
232828	Ruminococcaceae (Family)	Unknown Genus and Species	saccharin	6.31E-08	0.262	0.048
Collapsed	Clostridiales (Order)	Christensenellaceae (Family)	isoleucine	6.40E-08	0.074	0.014
189548	Lachnospiraceae (Family)	Unknown Genus and Species	cholesterol	6.66E-08	0.196	0.036
193763	Clostridiales (Order)	Unknown Family, Genus and Species	palmitate	6.70E-08	-0.314	0.058
199430	Ruminococcaceae (Family)	Faecalibacterium prausnitzii	palmitate	6.79E-08	-0.242	0.044
187917	Ruminococcaceae (Family)	Unknown Genus and Species	palmitate	7.16E-08	0.164	0.030
196724	Lachnospiraceae (Family)	Blautia (Unknown Species)	palmitate	7.39E-08	-0.215	0.039
208739	Ruminococcaceae (Family)	Faecalibacterium prausnitzii	palmitate	7.50E-08	-0.167	0.031
195494	Ruminococcaceae (Family)	Unknown Genus and Species	palmitate	8.22E-08	0.322	0.059
180462	Ruminococcaceae (Family)	Unknown Genus and Species	palmitate	8.46E-08	-0.206	0.038
Collapsed	Christensenellaceae (Family)	Unknown Genus	isoleucine	8.82E-08	0.072	0.013
194488	Ruminococcaceae (Family)	Unknown Genus and Species	palmitate	9.22E-08	0.277	0.051
4332082	Lachnospiraceae (Family)	Roseburia (Unknown Species)	palmitate	9.70E-08	-0.272	0.051
4383953	Clostridiaceae (Family)	Unknown Genus and Species	phenylacetate	9.92E-08	-0.082	0.015
183698	Lachnospiraceae (Family)	Unknown Genus and Species	phenylalanine	1.03E-07	-0.225	0.042
Collapsed	Clostridiales (Order)	Unknown Family	4-acetamidobutanoate	1.05E-07	0.045	0.008
Collapsed	Clostridiales (Order)	Unknown Family and Genus	4-acetamidobutanoate	1.05E-07	0.045	0.008
176062	Clostridiales (Order)	Unknown Family, Genus and Species	cholesterol	1.07E-07	0.075	0.014

Table A5: 297 nominally significant associations between adiposity and faecal metabolites.

Metabolite	Pheno	P Value	Estimate	SE
N-acetylisoleucine	VFM	0.0003	0.3379	0.0826
N-6-trimethyllysine	VFM	0.0003	0.3238	0.083
X-21742	BMI	0.0003	-0.7134	0.1726
Pseudouridine	VFM	0.0008	0.3232	0.0905
S-adenosylmethionine (SAM)	VFM	0.0009	0.3335	0.0943
N-acetyltyrosine	VFM	0.001	0.313	0.0896
N-acetyltyrosine	pTF	0.0011	0.327	0.0941
X-12095	pTF	0.0012	0.3293	0.0947
X-13696	pTF	0.0012	-0.3438	0.099
X-12700	VFM	0.0012	0.2777	0.0814
X-11612	VFM	0.0013	0.2961	0.087
2-pentanamido-3-phenylpropanoic acid	VFM	0.0016	0.3259	0.0968
2-aminoheptanoate	VFM	0.0019	0.2697	0.0824
X-17438	VFM	0.0019	-0.2842	0.0867
2-hydroxyoctanoate	pTF	0.0026	0.312	0.0897
5-methylthioadenosine(MTA)	VFM	0.0027	0.2755	0.0876
uracil	VFM	0.0029	0.2635	0.0842
5-methylthioadenosine(MTA)	pTF	0.003	0.2879	0.0917
lysylleucine	VFM	0.003	0.2884	0.089
5-methyluridine(ribothymidine.	VFM	0.0032	0.3038	0.097
cyclo(leu-phe)	VFM	0.0034	-0.3095	0.0980
X-21668	pTF	0.0034	0.3568	0.1147
phenyllactate(PLA)	VFM	0.0034	0.2715	0.0844

2-aminoadipate	pTF	0.0037	-0.2584	0.0798
X-15646	VFM	0.0037	0.3132	0.1030
4-hydroxycyclohexylcarboxylic acid	VFM	0.0041	0.2728	0.0901
(R)-salsolinol	pTF	0.0042	-0.3139	0.1035
allo-threonine	VFM	0.0045	0.2660	0.0896
3,4-hydroxyphenyl.lactate	VFM	0.0047	0.2585	0.0857
X-17009	VFM	0.0050	0.2712	0.0917
glycerophosphoethanolamine	pTF	0.0050	-0.3092	0.0779
X-14220	VFM	0.0051	0.2567	0.0857
X-12680	VFM	0.0051	0.2408	0.0826
guanosine	VFM	0.0052	0.2745	0.0926
uridine	pTF	0.0054	0.2687	0.0911
isoleucine	VFM	0.0059	0.2444	0.0850
inosine	VFM	0.0060	0.2477	0.0851
thymine	VFM	0.0061	0.2418	0.0845
glycylproline	VFM	0.0064	0.2387	0.0839
X-15192	VFM	0.0065	0.2216	0.0780
levulinate (4-oxovalerate)	pTF	0.0066	0.2561	0.0892
succinimide	VFM	0.0067	0.2371	0.0826
guanine	VFM	0.0071	0.2358	0.0827
X-23105	pTF	0.0071	-0.2650	0.0931
N-acetylleucine	VFM	0.0071	0.2485	0.0878
5-hydroxylysine	VFM	0.0073	0.2902	0.1022
inosine	pTF	0.0078	0.2535	0.0900
N-acetylglycine	BMI	0.0078	0.3632	0.1265
tyrosine	VFM	0.0079	0.2321	0.0843

methionine	VFM	0.0081	0.2293	0.0835
X-21353	BMI	0.0081	0.3048	0.1108
valine	VFM	0.0082	0.2304	0.0831
5alpha-pregnan-3beta,20alpha-diol monosulfate (2)	VFM	0.0085	0.2649	0.0963
X-13696	VFM	0.0085	-0.2716	0.0976
cadaverine	VFM	0.0086	0.2495	0.0915
bilirubin	BMI	0.0087	0.3488	0.1282
phenylalanine	VFM	0.0087	0.2302	0.0846
lysine	VFM	0.0088	0.2359	0.0860
histidylproline	VFM	0.0090	0.2331	0.0862
N-formylmethionine	VFM	0.0094	0.2244	0.0826
guanosine	pTF	0.0096	0.2698	0.0986
X-17842	VFM	0.0096	0.2680	0.1000
betA)alanine	VFM	0.0098	0.2198	0.0810
N-alphA)acetylornithine	VFM	0.0099	0.2477	0.0927
X-12193	BMI	0.0102	-0.3158	0.1095
N-acetylphenylalanine	VFM	0.0102	0.2253	0.0848
tyrosol	VFM	0.0102	0.2321	0.0872
leucine	VFM	0.0103	0.2200	0.0829
histamine	VFM	0.0104	0.2250	0.0846
2-pentanamido-3-phenylpropanoic acid	pTF	0.0105	0.2816	0.1057
S-adenosylmethionine (SAM)	pTF	0.0105	0.2810	0.1040
X-14264	pTF	0.0108	0.2508	0.0952
homoserine	VFM	0.0109	0.2227	0.0843
maltopentaose	pTF	0.0110	0.3259	0.1226
X-21756	VFM	0.0110	0.2309	0.0860

3-aminoisobutyrate	VFM	0.0112	0.2933	0.1117
spermidine	VFM	0.0113	0.2155	0.0822
cis-4-hydroxycyclohexylacetic acid	VFM	0.0117	0.2361	0.0887
alanylalanine	VFM	0.0121	0.2087	0.0805
acisoga	pTF	0.0123	0.2432	0.0921
X-21664	VFM	0.0124	0.2322	0.0882
carnosine	VFM	0.0125	0.2555	0.0977
succinate	VFM	0.0126	0.2119	0.0814
luteolin	pTF	0.0127	-0.2564	0.0992
xanthine	VFM	0.0129	0.2229	0.0855
nicotinate ribonucleoside	VFM	0.0130	0.2278	0.0869
2-ethylhexanoate	pTF	0.0131	0.2245	0.0861
N-acetylhistidine	VFM	0.0131	0.2166	0.0831
ornithine	VFM	0.0131	0.2263	0.0879
asparagylvaline	VFM	0.0131	0.2304	0.0891
X-22317	pTF	0.0134	0.2317	0.0889
erythritol	VFM	0.0134	0.2313	0.0896
asparagylisoleucine	VFM	0.0134	0.2377	0.0869
homoserine	pTF	0.0142	0.2274	0.0879
palmitoyl ethanolamide	pTF	0.0144	0.2304	0.0893
nicotinate ribonucleoside	pTF	0.0146	0.2362	0.0917
xanthosine	pTF	0.0147	0.2160	0.0839
pelargonate	BMI	0.0148	0.3006	0.1024
glycine	VFM	0.0148	0.2276	0.0906
2-pentanamido-3-phenylpropanoic acid	BMI	0.0149	0.3371	0.1333
nicotinate	VFM	0.0151	0.2066	0.0814

X-22317	VFM	0.0156	0.2142	0.0860
X-14697	VFM	0.0157	0.2089	0.0835
cysteine	VFM	0.0158	0.2273	0.0892
caprylate	pTF	0.0158	0.2285	0.0868
4-hydroxyphenylacetate	pTF	0.0159	0.2241	0.0881
erythronate.	VFM	0.0159	0.2135	0.0850
X-19751	pTF	0.0160	0.2371	0.0934
N-acetylmethionine	VFM	0.0160	0.2095	0.0844
X-20197	VFM	0.0161	0.2098	0.0844
hesperetin	VFM	0.0165	-0.2585	0.1039
3-hydroxypropanoate	pTF	0.0165	0.2083	0.0824
cadaverine	pTF	0.0170	0.2430	0.0967
alpha-glutamylvaline	VFM	0.0170	0.2146	0.0870
5-hydroxylysine	BMI	0.0174	0.3776	0.1542
gallate	pTF	0.0174	0.2397	0.0956
hypoxanthine	VFM	0.0175	0.2031	0.0829
histidine	VFM	0.0179	0.2090	0.0850
maltotetraose	pTF	0.0180	0.2382	0.0973
6-hydroxynicotinate	VFM	0.0183	0.2061	0.0833
oleic ethanolamide	pTF	0.0185	0.2287	0.0921
N2-acetyllysine	VFM	0.0186	0.2224	0.0904
X-22142	pTF	0.0191	0.2185	0.0885
X-14255	VFM	0.0192	0.2008	0.0834
2-aminoheptanoate	pTF	0.0195	0.2187	0.0905
X-12026	VFM	0.0196	0.2385	0.0992
5-hydroxy-2-indolecarboxylic acid	VFM	0.0197	-0.2886	0.1188

leucylglutamate	VFM	0.0198	0.2259	0.0940
indole	VFM	0.0198	0.2182	0.0897
glutamate	VFM	0.0198	0.2172	0.0884
palmitoyl sphingomyelin	BMI	0.0200	0.2661	0.1092
levulinate(4-oxovalerate)	VFM	0.0200	0.2078	0.0866
succinimide	pTF	0.0202	0.2170	0.0887
X-14502	pTF	0.0203	0.2310	0.0956
butyrylglycine	VFM	0.0203	0.2637	0.1077
X-12398	VFM	0.0206	0.2754	0.1153
X-11429	VFM	0.0206	0.2057	0.0850
naringenin	VFM	0.0207	-0.2307	0.0960
2-hydroxyadipate	VFM	0.0207	0.1999	0.0825
X-11522	BMI	0.0211	0.3084	0.1285
fucose	VFM	0.0214	0.2199	0.0906
X-19197	VFM	0.0218	0.2216	0.0916
valylglutamine	pTF	0.0218	0.2088	0.0886
X-14153	VFM	0.0219	0.2217	0.0917
2-aminobutyrate	VFM	0.0221	0.2122	0.0900
valylglutamate	VFM	0.0222	0.1992	0.0847
X-12125	VFM	0.0223	0.2269	0.0963
pyroglutamylglutamine	BMI	0.0224	-0.2704	0.1144
N-propionylmethionine	VFM	0.0226	0.2032	0.0867
X-02249	VFM	0.0227	0.1967	0.0837
stearidonate(18.4n3)	BMI	0.0228	0.2696	0.1114
threonine	VFM	0.0228	0.1988	0.0842
valylalanine	pTF	0.0228	0.2142	0.0916

6-acetyllysine	VFM	0.0229	0.1939	0.0820
threonylisoleucine	VFM	0.0230	0.2192	0.0891
X-11530	BMI	0.0231	0.3286	0.1385
2-deoxyinosine	VFM	0.0234	0.1985	0.0831
N-acetylvaline	VFM	0.0238	0.2161	0.0917
X-11612	pTF	0.0240	0.2278	0.0958
citrulline	VFM	0.0242	0.2134	0.0921
cysteine	BMI	0.0242	0.3174	0.1353
N-acetylputrescine	VFM	0.0242	0.1911	0.0814
isoleucylvaline	VFM	0.0245	0.2158	0.0901
glucosamine	BMI	0.0247	0.2736	0.1185
acetylcarnitine	pTF	0.0249	-0.2006	0.0869
X-21892	pTF	0.0250	0.2090	0.0837
ursodeoxycholate	BMI	0.0250	0.2870	0.1215
X1.stearoylglycerophosphoethanolamine	VFM	0.0251	-0.2526	0.0867
X-15192	pTF	0.0253	0.1961	0.0832
2-hydroxyoctanoate	VFM	0.0255	0.2127	0.0925
synephrine	VFM	0.0259	-0.2405	0.1042
pantothenate	VFM	0.0259	0.1875	0.0798
X-22167	VFM	0.0261	0.2416	0.1041
beta-alanine	pTF	0.0262	0.2022	0.0863
X-22146	VFM	0.0263	0.2163	0.0944
X-13891	VFM	0.0266	0.2105	0.0899
urate	VFM	0.0267	0.2230	0.0981
serylproline	VFM	0.0268	0.1851	0.0813
threonate	VFM	0.0273	0.1953	0.0860

sarcosine(N-Methylglycine)	pTF	0.0273	0.2444	0.1051
alpha-glutamylthreonine	VFM	0.0278	0.1995	0.0860
fructose	BMI	0.0278	-0.3053	0.1353
isoleucylglycine	VFM	0.0281	0.1993	0.0875
homocitrulline	VFM	0.0281	0.1859	0.0819
flavin adenine dinucleotide(FAD)	VFM	0.0282	0.2037	0.0880
N-acetylglutamine	pTF	0.0282	0.2092	0.0905
laurate(12-0)	pTF	0.0283	0.2072	0.0908
dimethylarginine	pTF	0.0283	0.2031	0.0879
dihydoxyphenylalanine	pTF	0.0287	-0.2056	0.0892
X-14272	VFM	0.0288	0.2181	0.0954
8.hydroxyguanine	VFM	0.0291	0.1919	0.0842
5-hydroxy-2-indolecarboxylic acid	pTF	0.0292	-0.2827	0.1262
citramalate	pTF	0.0294	0.2178	0.0949
N-acetylarginine	VFM	0.0301	0.1887	0.0825
X-14188	VFM	0.0303	0.2134	0.0935
O-sulfo-L-tyrosine	VFM	0.0303	0.2101	0.0934
valyllysine	VFM	0.0305	0.1872	0.0833
3-methylxanthine	pTF	0.0309	0.2272	0.0999
biocytin	VFM	0.0311	0.1932	0.0846
indolin-2-one	BMI	0.0314	-0.2789	0.1266
X-14155	VFM	0.0315	0.2065	0.0933
spermidine	pTF	0.0315	0.1963	0.0866
phenylalanylalanine	pTF	0.0315	0.1998	0.0906
beta-hydroxyisovalerate	VFM	0.0316	0.2039	0.0888
valylaspartate	VFM	0.0320	0.1920	0.0869

eriodictyol	pTF	0.0323	-0.2410	0.1094
X-14264	VFM	0.0323	0.2024	0.0923
xanthosine	VFM	0.0323	0.1825	0.0808
valylserine	pTF	0.0324	0.2004	0.0910
maltotriose	BMI	0.0324	-0.2688	0.1222
X-14220	BMI	0.0324	0.2529	0.1154
alanine	VFM	0.0326	0.1817	0.0829
X-16975	VFM	0.0327	0.2071	0.0942
tyrosylvaline	pTF	0.0330	0.1939	0.0888
lysylvaline	pTF	0.0330	0.1971	0.0901
lysylphenylalanine	VFM	0.0331	0.1972	0.0853
X-15856	pTF	0.0332	-0.1922	0.0857
uridine	VFM	0.0335	0.2004	0.0894
X-14263	VFM	0.0336	0.1950	0.0896
hesperetin	BMI	0.0338	-0.2841	0.1266
putrescine	VFM	0.0341	0.1863	0.0854
X-16581	pTF	0.0342	0.2492	0.1116
X-13735	VFM	0.0342	0.2122	0.0973
X-15136	pTF	0.0345	-0.1828	0.0837
X-17327	VFM	0.0346	0.2114	0.0977
methionylglutamate	VFM	0.0350	0.1886	0.0871
7-ketodeoxycholate	BMI	0.0351	0.2287	0.1052
N-acetylvaline	BMI	0.0352	0.2561	0.1183
X-14155	BMI	0.0356	0.2580	0.1200
arginylisoleucine	pTF	0.0359	0.2083	0.0951
X-12734	pTF	0.0361	0.2262	0.1031

threonylglutamate	VFM	0.0365	0.1961	0.0894
isoleucylleucine	pTF	0.0367	0.1788	0.0834
N-acetylisoleucine	pTF	0.0374	0.2069	0.0961
norvaline	VFM	0.0374	0.1944	0.0890
ponciretin	VFM	0.0374	-0.2472	0.1161
threonylproline	VFM	0.0376	0.1840	0.0862
methionine sulfoxide	VFM	0.0377	0.1902	0.0892
carboxyethyl(GABA)	pTF	0.0378	0.1976	0.0903
isoleucylglutamine	pTF	0.0384	0.1842	0.0870
N-acetylglutamine	VFM	0.0385	0.1872	0.0867
alpha-glutamyltyrosine	VFM	0.0393	0.1833	0.0860
X-21892	VFM	0.0393	0.1759	0.0835
catechol	pTF	0.0395	0.1996	0.0920
lysylvaline	VFM	0.0401	0.1889	0.0870
7-methylxanthine	pTF	0.0401	0.2124	0.0981
tryptophan	VFM	0.0401	0.1863	0.0888
X-21756	pTF	0.0401	0.1980	0.0926
N1-methylguanosine	pTF	0.0402	0.1994	0.0922
X-19375	pTF	0.0404	0.2155	0.1020
N-acetylglucosamine	VFM	0.0410	0.1870	0.0872
X-11429	pTF	0.0414	0.1943	0.0903
ethanolamine	VFM	0.0416	0.1726	0.0803
proline	VFM	0.0418	0.1866	0.0892
3-hydroxyindolin-2-one	VFM	0.0418	0.1832	0.0866
inositol-1-phosphate	BMI	0.0419	0.2204	0.1038
ribonate	VFM	0.0420	0.1870	0.0881

X-23105	VFM	0.0422	-0.1946	0.0926
X-14894	BMI	0.0427	-0.2097	0.1003
N-delta-acetylornithine	VFM	0.0433	0.1755	0.0848
caffeate	BMI	0.0433	-0.2637	0.1275
5-methyluridine	pTF	0.0439	0.2273	0.1069
trans-4-hydroxyproline	VFM	0.0439	0.1911	0.0923
X-20100	VFM	0.0440	0.1730	0.0833
N-acetylglycine	VFM	0.0443	0.2201	0.1048
tryptophylproline	VFM	0.0443	0.1706	0.0827
alanylproline	VFM	0.0445	0.1775	0.0862
docosapentaenoate	BMI	0.0446	0.2342	0.1128
N-acetylglucosamine	pTF	0.0449	0.1941	0.0917
5-acetylamino-6-formylamino-3-methyluracil	pTF	0.0450	0.2035	0.0965
X-21757	VFM	0.0450	0.1724	0.0834
4-hydroxycyclohexylcarboxylic acid	pTF	0.0451	0.2092	0.0990
lysylisoleucine	VFM	0.0454	0.1834	0.0860
X-15863	VFM	0.0455	0.1927	0.0943
5-ketogluconate	pTF	0.0456	0.1913	0.0907
X-19585	pTF	0.0457	-0.1811	0.0859
glycylphenylalanine	VFM	0.0458	0.1843	0.0890
isoleucylmethionine	VFM	0.0458	0.1764	0.0840
tyrosylglycine	pTF	0.0462	0.1956	0.0930
X-17327	pTF	0.0466	0.2103	0.1031
glycyltryptophan	VFM	0.0468	0.1822	0.0898
tyrosylglutamate	pTF	0.0469	0.1899	0.0906
N-N-dimethylsphingosine	pTF	0.0470	0.1836	0.0906

theanine	pTF	0.0471	0.2006	0.0990
thymidine	VFM	0.0472	0.1754	0.0839
cholesterol	pTF	0.0475	0.1839	0.0879
phenylalanylproline	VFM	0.0476	0.1744	0.0863
X-21471	pTF	0.0479	0.2478	0.1187
hyocholate	pTF	0.0482	0.1872	0.0925
alpha-hydroxyisovalerate	BMI	0.0485	0.2221	0.1102
succinate	pTF	0.0486	0.1820	0.0874
phenylalanylglutamate	pTF	0.0488	0.1877	0.0932
X-18034	VFM	0.0490	0.1725	0.0850
5alpha-pregnan-3beta-20alpha-diol-monosulfate	pTF	0.0491	0.2095	0.1033
threonylvaline	VFM	0.0495	0.1801	0.0895
5-aminovalerate	BMI	0.0498	0.2368	0.1181
tryptophylglutamate	pTF	0.0500	0.1976	0.0955

B

APPENDIX B: PUBLISHED PAPERS AND PAPERS IN DEVELOPMENT

B.1 APPENDIX B1. HERITABLE COMPONENTS OF THE HUMAN GUT MICROBIOME ARE ASSOCIATED WITH VISCERAL FAT - IN PREPARATION

The manuscript in this section is an update of the analyses presented in Chapter 5. The dataset was expanded to include a total of 1313 individuals with obesity and 16S data. This manuscript is currently in draft form and I am performing ongoing analyses to include in the final draft.

B.1 APPENDIX B1. HERITABLE COMPONENTS OF THE HUMAN GUT MICROBIOME ARE ASSOCIATED WITH VISCERAL FAT - IN PREPARATION

Heritable components of the human gut microbiome are associated with visceral fat

Michelle Beaumont[1], Julia K. Goodrich[2,3], Matthew A. Jackson[1], Idil Yet[1], Emily Davenport[3], Tess Pallister[1], Justine Debelius[4], Rob Knight[4,5,6,7], Andrew G. Clark[2], Ruth E. Ley[2,3], Tim D. Spector[1]*, Jordana T. Bell[1]*

[1]Dept of Twin Research & Genetic Epidemiology, King's College London, U.K.
[2]Dept of Microbiology, Cornell University, Ithaca, NY 14853 US
[3]Dept of Molecular Biology & Genetics, Cornell University, Ithaca, NY 14853, USA
[4]Dept of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA
[5]Biofrontiers Institute, University of Colorado, Boulder, CO 80309, USA
[6]Howard Hughes Medical Institute, Boulder, CO 80309, USA
[7]Current address: Departments of Pediatrics and Computer Science & Engineering, University of California San Diego, La Jolla CA 92093, USA

* Equal contribution

Corresponding:

Jordana T. Bell
Department of Twin Research & Genetic Epidemiology
King's College London
St Thomas' Hospital
3rd Floor, South Wing, Block D
London
United Kingdom
SE1 7EH
Tel: (+44) 0207 18 81506
Email: jordana.bell@kcl.ac.uk

Tim D. Spector
Department of Twin Research & Genetic Epidemiology
King's College London
St Thomas' Hospital
3rd Floor, South Wing, Block D
London
United Kingdom
SE1 7EH
Tel: (+44) 0207 18 86739
Email: tim.spector@kcl.ac.uk

Abstract

The human gut microbiota has previously been associated with body mass index (BMI). Although obesity is a global health burden it is specifically the accumulation of abdominal fat that is a crucial cardio-metabolic disease risk factor. We explored links between the gut microbiota and abdominal adiposity in 1,313 twins, comparing 16S rRNA profiles with visceral fat, subcutaneous fat, trunk fat, waist-to-hip ratio, android-gynoid ratio, and BMI. We identified 149 associations between 97 gut microbes and abdominal adiposity measures with the strongest associations between visceral fat and Oscillospira members (minimum P = 1.93×10^{-12}). Heritability of visceral fat (estimated at $h^2 = 0.73$) and the adiposity-associated microbial units (average estimated $h^2 = 0.16$) prompted us to assess host genetic-microbe interactions at obesity-associated human candidate loci. We observed significant associations with host genetic variants in the *FHIT, TDRG1* and *ELAV* genes, suggesting a role for host genetic influence over the link between the gut microbiome and obesity. Our results give insights into the role of the gut microbiota in cardio-metabolic disease and potential for health management.

Introduction

Obesity has rapidly become a global public health problem with obesity-related disease now one of the leading causes of preventable death worldwide[1]. Although overall obesity poses a global health epidemic, it is the accumulation of excess abdominal fat that is a critical risk factor for cardiovascular and metabolic disease. Changes in diet and sedentary lifestyle can partly explain obesity, but family and twin studies also show a genetic influence, with obesity heritability estimates of 60-70%[2-5]. Genome-wide association studies (GWAS) have identified genetic risk factors[6-8], but genetic variants detected to date explain less than 3% of the heritability of obesity, with prediction ability of up to 20%, suggesting a role for other mechanisms[9].

Recent insights show that the gut microbiota may play a crucial role in obesity and cardio-metabolic disease risk. Many studies have linked different aspects of the gut microbiome to obesity[10-15]. However, the causal mechanisms leading to these associations are unclear and the bacteria identified differ between studies. Inconsistencies in the taxa associated with obesity may be in part explained by study design differences such as control of diet and sequencing platforms, but could also be due to differences in collective bacterial gene function rather than the species community composition[16,15]. A recent meta-analysis has shown that results between obese microbiome studies are not consistent and there is no clear Firmicutes/Bacteroides ratio change as is commonly posited[15]. Another complicating factor is phenotype measurement. While most human studies consider BMI as the measure of obesity[10,17,18], mouse studies typically use epididymal fat weight[19] or Dual-Energy X-ray Absorptiometry (DEXA)-derived measures of total body fat[20,21]. BMI is imprecise and measures overall adiposity without distinction between lean and fat mass[22]. Estimates of visceral fat, however, have stronger associations with obesity-related cardio-metabolic diseases, such as type 2 diabetes and cardiovascular disease[23-25], but have typically been difficult to measure in humans.

A recent study by Goodrich et al (2014)[26] was the first to report heritability of the human gut microbiota, with the abundance of the family Christensenellaceae showing the most variance attributed to host genetic effects (39%). Christensenellaceae was also enriched in abundance in the microbiomes of low-BMI individuals. Here, we build upon these findings to explore the association between the human gut microbiome and abdominal adiposity, as the main risk factor for cardio-metabolic disease risk. We obtained DEXA-based measures of abdominal adiposity, specifically, visceral fat mass and previously reported trunk fat measures[27] in the twin sample profiled by Goodrich et al (2014)[26]. We

show that heritable components of the human gut microbiome[26] are significantly associated with visceral fat, confirming the key role of the microbiome in cardio-metabolic disease risk. We further identify a link between gut microbiome profiles, visceral fat and trunk fat, and genetic variants in the human obesity candidate gene *FHIT*, providing a potential mechanism relating the gut microbiome to cardio-metabolic disease risk.

Results

Measures of adiposity were obtained from 1,313 predominantly female twins from the TwinsUK cohort (496 monozygotic (MZ), 592 dizygotic (DZ), 2 with unknown zygosity and 223 unrelated individuals; average age 63 years (range 32-87); 96.4% female). We obtained gut microbiome profiles in the 1,313 individuals by sequencing faecal samples from 353 individuals and combining these data with 960 previously published profiles [26]. All faecal samples underwent 16S rRNA profiling on the Illumina MiSeq platform, providing 2,135 operational taxonomic units (OTUs) at 97% sequence identity.

Adiposity and Visceral fat heritability

We studied 6 adiposity measures in total, and these included 3 measures of abdominal adiposity (visceral fat mass (VFM), subcutaneous fat mass (SFM), percentage trunk fat (pTF)), 2 measures of body fat distribution (android/gynoid ratio (AGR) and waist-hip ratio (WHR)), and one measure of overall obesity, BMI. Adiposity was estimated using DEXA-derived measures, which have been shown to be reliable alternatives[28-31] to traditional CT and MRI scan-based measures of adiposity. The majority of these adiposity measures have been explored in the TwinsUK prior to this study, however visceral fat is a novel phenotype in this cohort. The new measure of VFM was highly correlated with other abdominal and overall adiposity measures, including BMI. Twin-based heritability analysis of VFM showed evidence of a significant additive genetic component, or narrow-sense heritability (h^2), contributing to 0.73 (95% CI = 0.62 – 0.78) of the variance. We obtained comparable estimate for the narrow-sense heritability of BMI (h^2=0.76 (95% CI = 0.53 – 0.81) and a lower estimate for pTF (h^2=0.57 (95% CI = 0.50 – 0.64), in line with previous studies[32-34].

The twin gut microbiome and its heritability

Because our sample contained twins, we next explored evidence for heritability in the gut microbial profiles, extending the results of Goodrich et al (2014)[26] in the larger sample of 1,313 twins. Using the methods outlined in Goodrich et al (2014)[26] we calculated heritability for all OTUs in the larger dataset. Altogether, OTU narrow-sense heritability in this dataset ranged between 0 and 0.42, and the average estimate over all OTUs was 0.06. The most heritable microbe was *Clostridum perfringens* (h^2=0.42 (95%CI 0.23-0.51)). *Christensenella* was the most heritable microbe reported in Goodrich et al (2014)[26], and while it was not the most heritable ranked in this larger dataset, it remained highly heritable with a heritability of 0.31 (95%CI 0.21-0.41). Approximately 6% of the dataset (X OTUs) had evidence for at least moderate heritability (h^2>0.2).

Gut microbiome diversity is strongly linked to obesity and central adiposity

Microbial diversity in obese individuals has been reported to be lower than that of lean individuals[35]. Here we estimated alpha diversity of the sample using phylogeny-based metrics, and report Shannon diversity. We compared Shannon Diversity with all adiposity measures using a linear mixed effects model, adjusting for dietary principle components (see Methods), age, sex and family relatedness. As in previous literature, we observed a significant negative association between Shannon diversity and all adiposity phenotype. VFM showed the most significant association (beta = , se = , P=4.13x10^{-7}) and pTF showed the least significant association, though it surpassed nominal significance (P=1.51x10^{-2}). Therefore, alpha diversity measures in our sample are not only negatively associated with obesity, but are also significantly lower in individuals with greater abdominal adiposity and visceral fat.

Gut microbiome profiles associate with central adiposity across twins

We investigated the association of each OTU and collapsed taxonomic trait with all adiposity traits, including BMI, across individuals.

OTUs found in less than 25% of individuals were discarded and the remaining counts converted to relative abundances. Residuals from models including technical covariates such as sequencing batch and depth of sample sequencing were used in analyses to control for technical and batch effects.

Of the approximately 12,000 OTU-phenotype associations considered, 3,217 were nominally significant and 140 OTU results surpassed Bonferroni correction ($P = 3.90x10^{-6}$). Of the 140 significant microbial-adiposity associations, there were 94 unique OTUs, and these fell within either the Firmicutes or Bacteroidetes phyla, and most within the Ruminococcaceae family Visceral fat (VFM) associations made up the highest proportion of results surpassing the Bonferroni threshold (45%), and comprised the most significant associations. The peak result was an OTU classified as *Oscillospira* (Greengenes OTU 352014), which was associated with VFM (ANOVA $p = 1.91x10^{-11}$). Ruminococcaceae OTUs featured prominently in the top significant results, along with a number of other OTUs within the Lachnospiraceae family. These results support the crucial role of the microbiome towards visceral fat as a marker of adiposity and cardio-metabolic disease risk.

We also considered gut microbiome differences within a subsample of MZ twins, because MZ twins are genetically identical to each other and the differences observed between such twins are likely to be a result of environmental or epigenetic influences. In order to explore this, we estimated gut microbiome differences within MZ twins and compared these to VFM differences within twin pairs, but the results did not surpass correction for multiple testing.

Because of the strong association that we observed between alpha diversity and adiposity, we also wanted to assess if the strongest adiposity-OTU associations were simply for OTUs that were markers of diversity, or whether these taxa associated with adiposity were independent of species richness. To this end we repeated the adiposity-OTU analyses at the 149 significant OTU-phenotype pairs but now including alpha diversity as a covariate in the liner model. All of the reported significant associations with VFM and BMI remained nominally significant after adjustment for alpha diversity, but several associations with pTF dropped below the threshold for statistical significance. Overall, most associations with all three phenotypes are independent of alpha diversity.

Replication of BMI association in American Gut

We pursued replication of the significant OTUs associated with BMI in an independent sample of 1971 individuals from the American Gut (AG) cohort. We initially compared our OTU profiles to those generated in the AG data.

Individuals were selected from the American Gut data set with complete technical and biological covariate data, including sequencing depth and plate, smoking status, gender, bmi and age. Unrarefied OTU counts from this set were then transformed to relative abundances as for TwinsUK OTUs, again discarding those in <25% individuals. The ability of mixed effects models with and without BMI was then assessed by Anova to identify BMI associated OTU within American Gut.

Of the 97 unique OTUs that comprised the 149 significant adiposity associations in the TwinsUK sample, 16 OTUs were also found within the AG dataset. All of the 16 OTUs were at least nominally significantly associated with BMI in the TwinsUK sample. We were able to replicate the BMI-OTU associations at 8 of the 16 OTUs within the AG dataset, at nominal significance and in the same direction of association. We also considered all OTUs that showed evidence for association with BMI in the TwinsUK cohort, and found that 101 of these are also present in the AG dataset. Of the 101, 14 OTUs replicated at nominal significance with the same direction of effect, and a further 25 OTUs had the same direction of effect, but were not nominally significant.

Host genetic influences on microbiome-obesity associations

Because our dataset consisted of a familial sample of twins, we could dissect the contribution of genetic and environmental components to both estimates of the gut microbiome and abdominal adiposity. Twin-based heritability estimates supported a strong genetic component for visceral fat and our heritability analyses of the gut microbiome in this sample showed wide variability in heritability between taxa (0–0.42), with members of Firmicutes and Actinobacteria being most heritable. The peak Bonferroni significant OTUs associated with abdominal adiposity in our sample had an average OTU heritability of 0.16, suggesting potential host genetic links to the identified adiposity associations.

To explore the hypothesis that host genetics may influence the observed microbial-adiposity associations, we performed candidate gene analysis comparing host genetic variants at human obesity candidate loci with the adiposity-associated gut microbiome profiles. We selected SNPs within human loci previously associated with obesity as reported by Locke et al (2015)[7], using common genetic variants within 97 50kb regions centred around peak BMI-associated GWAS SNPs. At a Bonferroni-corrected P-value threshold taking into account the total number of genomic regions and adiposity-OTUs considered (P = 5.31e-6), OTU associations with genetic variants in 3 genomic regions surpassed multiple testing. The strongest association between host genotype and adiposity-associated OTUs were observed between a OTU within the Clostridiales order (Greengenes OTU 181702) and a host genetic

variant within an intron of the *FHIT* gene (rs74331972 with OTU 181702 P = 2.49E-06). *FHIT* encodes the fragile histidine triad protein and is a tumour suppressor gene that has been linked to cancers of the digestive tract.

When we considered the genetic-OTU association results at a more relaxed significance threshold (P = $5x10^{-8}$), there were a total of 412 suggestive OTU-genetic associations located within or near 48 unique genes, including key obesity genes such as *FTO*, *RPTOR* and *TMEM18*.

Discussion

Visceral fat associates with Firmicutes members, Lachnospiraceae and Ruminococcaceae

Here, in the largest microbiota-phenotype study to date using detailed visceral fat measures, we have shown how specific members of the human gut microbiota clearly associate with obesity-related phenotypes, most notably those relating to VFM. Due to the importance of visceral fat in cardio-metabolic disease risk, direct measures of visceral fat are more informative than BMI in assessing the metabolic consequences of obesity on obesity-related disease. Most microbiome obesity studies to date have used BMI as a biomarker of obesity, with some mouse studies using epididymal fat. We therefore present a dual approach, using BMI to confirm associations and visceral fat to find novel metabolic associations in a sample of twins. The power of this approach was apparent in the number of top associations that were observed with visceral fat. The Firmicutes phylum showed the most significant associations, yet the direction of association differed for different genera within the Firmicutes. Furthermore, the majority of our top hits are independent of alpha diversity, suggesting that the OTUs significant in this analysis are not just markers of diversity, but form real associations. In the American Gut cohort, our replication sample, we observed that the same families, Lachnospiraceae and Ruminococacceae were associated with BMI. Firmicutes have previously been associated with obesity in some but not all studies, and several of these studies show an increase in Firmicutes in obese subjects[11,43,44]. For example, *Ruminococcus gnavus* has been shown to be significantly enriched in low microbial gene count individuals that were prone to obesity in one study[14]. However, the phylum

Firmicutes contains over 270 genera with many different and diverse functions. Because changes may occur at finer-grained taxonomic resolution, and because the baseline abundance of different genera differs among human populations, opposing selection processes at different phylogenetic levels may in part explain differing results in studies determining microbiota differences in obesity[45,46], including recent meta-analyses across multiple obesity microbiota datasets[15,47]. This difference in directional effects in genera of the same family has previously been observed in animal models. Mice consuming a Western diet have previously been reported to have increased levels of *Eubacterium dolichum*[48], while another study has shown that mice consuming a high-fat diet have decreased levels of *Allobaculum* OTUs[49]. Both of these microbes are members of the Erysipelotrichaceae family, and show opposite directions of effect.

Heritability of visceral fat

Visceral fat, the type of adipose tissue with the most important implications for metabolic health was highly heritable (73%) and showed the most significant associations with the gut microbiota. Previous studies, for example the Framingham[32], Quebec[34] and Heritage[33] cohorts have found heritability estimates of visceral fat to be between 36% and 55%, that is lower but comparable to our estimates in twins. However another study in the same twin cohort found visceral fat heritability to be approximately 58%[31]. This study used estimates of visceral fat that were derived and modelled using a combination of DEXA and anthropometric measures, potentially accounting for the lower heritability estimate as the visceral fat estimate is not measured as directly as in this study. This is the first time, to our knowledge, that visceral fat has been linked with changes in the gut microbiota in humans. The results suggested that fat mass is more important for differences in the obese microbiome, rather than overall body mass. This is also demonstrated in the recent Clarke et al (2014) study[50], where rugby union players, who due to their greater muscle mass were defined as overweight and obese (BMI = 29 ± 3), had a more diverse microbiota than that of both low BMI and high BMI controls, likely due to differences in exercise.

Conclusion

In conclusion, we present the novel finding that members of the Clostridiales associate significantly with abdominal visceral fat, which was highly heritable in our sample. Furthermore, we have identified a

promising gene candidate that may influence the interaction between the human gut microbiome and obesity and its metabolic consequences.

Methods

Sample collection

Faecal samples were collected from 1,313 healthy, volunteers from the TwinsUK Adult Twin Registry (496 MZ, 592 DZ, 2 with unknown zygosity and 223 unrelated individuals; 47 male, 1266 female). Samples were refrigerated for 1-2 days prior to the participant's annual clinic visit immediately at which they were stored for up to 8 weeks at -80°C before DNA extraction. The project was approved by the local ethics committee and signed and written consent was obtained from all participants.

Gut microbiome profiles

Needs Input.

960 16S gut microbiome profiles in the twin sample were previously derived by Goodrich et al[26]. Briefly, DNA extracted from faecal samples underwent amplification of the V4 region of the 16S rRNA gene, followed by paired-end sequencing on the Illumina MiSeq platform.

OTUs not found in at least 25% of individuals were discarded and their counts were converted to relative abundances within samples, followed by the addition of a pseudocount (10E-6) to remove zero counts. Mixed effects models were constructed including sequencing run, depth of sample sequencing, and technician who extracted and loaded the DNA for sequencing, as well as sample collection method (by post or in person) as covariates with OTU abundances as the response. Residuals from these models were then used in subsequent analyses to control for technical and batch effects.

Following sequence data processing and quality control (see Goodrich 2014[26]), and adjustment for age, gender, shipment, number of sequences per sample and sequencing run, the final dataset contained 2135 OTUs.

OTUs from the complete set (including those in <25% of individuals) were then collapsed into genera and families based on shared taxonomic assignment. Taxa found in less than 10 individuals were then discarded and the counts converted to relative abundances. Models were used to control for technical covariates as for OTUs and residuals used in subsequent analyses.

Microbiome profiles from the American Gut cohort were used as a replication sample. Subjects were consented under the protocol approved by the University of Colorado IRB. Samples were collected at home and mailed in to the lab. These were extracted using the Power Soil kit, amplified using the EMP primers and sequenced using MiSeq. OTUs were picked against Greengenes May 2013 release, and bloom filtered to computationally remove overabundant gammaproteobacterial OTUs using UClust in QIIME[53]. A subset of samples were selected where data was publicly available and was complete for sequencing plate, sequencing depth (taken as the ENA base count), smoking status, gender, age and BMI (n=1101). We explored replication specifically from 10 peak-associated OTUs within Lachnospiraceae and Ruminococcaceae, as well as all reported significant associations within BMI (Supplementary Table 1).

Phenotype data

Obesity phenotype data were collected during each participant's annual clinic visit. We explored 6 obesity-related phenotypes in this work for a subset of 1313 twins (496 MZ, 592 DZ, 2 with unknown zygosity and 223 unrelated individuals). These included two measures from total body DEXA (Dual-Energy X-ray Absorptiometry) whole-body scanning. During the DEXA procedure participants were asked to lie flat and straight while a full body scan took place, taking measures for percentage trunk fat and visceral fat mass (g)[27]. Visceral fat mass was calculated from one cross-section of the whole body at L4-L5, the typical location of a CT slice. In addition we also had BMI data available for all 1313 twins.

Statistical analysis

To assess the evidence for association between gut microbiome composition and obesity-related and metabolic phenotypes, we performed two main analyses. First, we compared microbiome OTUs and phenotypes by fitting a linear mixed effects model (LMER) using the R package LME4. In this model the phenotype (or metabolite where appropriate) was the response variable and the OTU was a fixed effect predictor, family and zygosity were taken into account as random effects and sex, diet principle

components and age were considered as fixed effects. Each phenotype was normalised to a standard Normal distribution prior to the analysis. To assess the significance of the association, we compared the full regression model described above, to a null model that excluded the OTU predictor. We considered all associations that surpassed nominal significance ($p < 0.05$) and passed an FDR 1%.

Replication was undertaken using 16S microbiome data from the American Gut cohort. American Gut data was subset to the individuals with complete data and the unrarefied OTU counts were transformed to relative abundances, discarding OTUs in <25% of individuals. The OTU abundances were used in turn as the response variables in mixed effects models including sequencing plate, sequencing depth, smoking status, gender and age as covariates, with sequencing plate as a random effect. The ability of models with and without the addition of BMI as a predictor was then assessed by Anova to identify BMI associated OTUs within American Gut.

Heritability was assessed using the ACE model. Under the assumption that the dominance effects are negligible, the ACE model can estimate the additive genetic (A), common environment (C) and unique environment (E) components of the trait variance. Twin heritability estimates were obtained using OpenMX[56].

Candidate gene analysis was performed using the software GEMMA (Genome-wide Efficient Mixed Model Association)[57] on 1059 individuals with genotypes. GEMMA implements a univariate linear mixed model to perform association tests, using a kinship matrix to take into account twin relatedness. The list of candidate obesity GWAS SNP associations was taken from Locke et al[ref]. The reported SNPs were taken as the lead SNP, and candidate gene regions were extended to include additional SNPs within a 25 kb region either side of the lead SNP. Overall, we considered associations with 1795 SNPs across 53 unique genomic regions. No MAF criteria were applied during the initial SNP selection.

Funding

This work was funded by NIH RO1 DK093595, DP2 OD007444, The Cornell Center for Comparative Population Genomics, the Wellcome Trust, and the European Community's Seventh Framework Programme (FP7/2007-2013). The study also received support from the National Institute for Health Research (NIHR) BioResource Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London.

Acknowledgements

We acknowledge the twins and TwinsUK for data access. TwinsUK was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013) and also receives support from the National Institute for Health Research (**NIHR**)- funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. SNP genotyping was performed by The Wellcome Trust Sanger Institute and National Eye Institute via NIH/CIDR**.**

Conflict Of Interest

The authors declare no conflict of interest.

References

1 Barness, L. A., Opitz, J. M. & Gilbert-Barness, E. Obesity: Genetic, molecular, and environmental aspects. *American Journal of Medical Genetics Part A* **143A**, 3016-3034, doi:10.1002/ajmg.a.32035 (2007).

2 Maes, H. H. M., Neale, M. C. & Eaves, L. J. Genetic and environmental factors in relative body weight and human adiposity. *Behavior Genetics* **27**, 325-351, doi:10.1023/a:1025635913927 (1997).

3 Allison, D. B. *et al.* The heritability of body mass index among an international sample of monozygotic twins reared apart. *International Journal of Obesity* **20**, 501-506 (1996).

4 O'Rahilly, S. & Farooqi, I. S. Human Obesity: A Heritable Neurobehavioral Disorder That Is Highly Sensitive to Environmental Conditions. *Diabetes* **57**, 2905-2910, doi:10.2337/db08-0210 (2008).

5 Samaras, K. *et al.* Independent genetic factors determine the amount and distribution of fat in women after the menopause. *Journal of Clinical Endocrinology & Metabolism* **82**, 781-785, doi:10.1210/jc.82.3.781 (1997).

6 Poirier, P. *et al.* Obesity and cardiovascular disease - Pathophysiology, evaluation, and effect of weight loss. *Arteriosclerosis Thrombosis and Vascular Biology* **26**, doi:10.1161/01.ATV.0000216787.85457.f3 (2006).

7 Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206, doi:10.1038/nature14177

http://www.nature.com/nature/journal/v518/n7538/abs/nature14177.html - supplementary-information (2015).

8 Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187-196, doi:10.1038/nature14132

http://www.nature.com/nature/journal/v518/n7538/abs/nature14132.html - supplementary-information (2015).

9 Yang, W., Kelly, T. & He, J. Genetic epidemiology of obesity. *Epidemiologic Reviews* **29**, doi:10.1093/epirev/mxm004 (2007).

10 Claesson, M. J. *et al.* Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**, doi:10.1038/nature11319 (2012).

11 Ley, R. E. *et al.* Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences of the United States of America* **102**, doi:10.1073/pnas.0504978102 (2005).

12 Armougom, F., Henry, M., Vialettes, B., Raccah, D. & Raoult, D. Monitoring Bacterial Community of Human Gut Microbiota Reveals an Increase in Lactobacillus in Obese Patients and Methanogens in Anorexic Patients. *Plos One* **4**, doi:10.1371/journal.pone.0007125 (2009).

13 Everard, A. *et al.* Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 9066-9071, doi:10.1073/pnas.1219451110 (2013).

14 Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541-+, doi:10.1038/nature12506 (2013).

15 Walters, W. A., Xu, Z. & Knight, R. Meta-analyses of human gut microbes associated with obesity and IBD. *Febs Letters* **588**, 4223-4233, doi:10.1016/j.febslet.2014.09.039 (2014).

16 Tilg, H. & Kaser, A. Gut microbiome, obesity, and metabolic dysfunction. *Journal of Clinical Investigation* **121**, doi:10.1172/jci58109 (2011).

17 Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457**, 480-484 (2009).

18 Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174-180, doi:10.1038/nature09944 (2011).

19 Vijay-Kumar, M. *et al.* Metabolic Syndrome and Altered Gut Microbiota in Mice Lacking Toll-Like Receptor 5. *Science* **328**, 228-231, doi:10.1126/science.1179721 (2010).

20 Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027-1031, doi:10.1038/nature05414 (2006).

21 Turnbaugh, P. J., Baeckhed, F., Fulton, L. & Gordon, J. I. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host & Microbe* **3**, 213-223, doi:10.1016/j.chom.2008.02.015 (2008).

22 Romero-Corral, A. *et al.* Accuracy of body mass index in diagnosing obesity in the adult general population. *International Journal of Obesity* **32**, 959-966, doi:10.1038/ijo.2008.11 (2008).

23 Wajchenberg, B. L. Subcutaneous and visceral adipose tissue: Their relation to the metabolic syndrome. *Endocrine Reviews* **21**, 697-738, doi:10.1210/er.21.6.697 (2000).

24 Fontana, L., Eagon, J. C., Trujillo, M. E., Scherer, P. E. & Klein, S. Visceral fat adipokine secretion is associated with systemic inflammation in obese humans. *Diabetes* **56**, 1010-1013, doi:10.2337/db06-1656 (2007).

25 Saito, T. *et al.* Association of subcutaneous and visceral fat mass with serum concentrations of adipokines in subjects with type 2 diabetes mellitus. *Endocrine Journal* **59**, 39-45 (2012).

26 Goodrich, Julia K. *et al.* Human Genetics Shape the Gut Microbiome. *Cell* **159**, 789-799, doi:http://dx.doi.org/10.1016/j.cell.2014.09.053 (2014).

27 Livshits, G., Kato, B., Wilson, S. & Spector, T. Linkage of genes to total lean body mass in normal women. *Journal of Clinical Endocrinology & Metabolism* **92**, 3171-3176, doi:10.1210/jc.2007-0418 (2007).

28 Snijder, M. B. *et al.* The prediction of visceral fat by dual-energy X-ray absorptiometry in the elderly: a comparison with computed tomography and anthropometry. *International Journal of Obesity* **26**, 984-993, doi:10.1038/sj.ijo.0801968 (2002).

29 Bertin, E., Marcus, C., Ruiz, J. C., Eschard, J. P. & Leutenegger, M. Measurement of visceral adipose tissue by DXA combined with anthropometry in obese humans. *International Journal of Obesity* **24**, 263-270, doi:10.1038/sj.ijo.0801121 (2000).

30 Kaul, S. *et al.* Dual-Energy X-Ray Absorptiometry for Quantification of Visceral Fat. *Obesity* **20**, 1313-1318, doi:10.1038/oby.2011.393 (2012).

31 Direk, K. *et al.* The relationship between DXA-based and anthropometric measures of visceral fat and morbidity in women. *Bmc Cardiovascular Disorders* **13**, doi:10.1186/1471-2261-13-25 (2013).

32 Fox, C. S. *et al.* Abdominal visceral and subcutaneous adipose tissue compartments - Association with metabolic risk factors in the Framingham Heart Study. *Circulation* **116**, 39-48, doi:10.1161/circulationaha.106.675355 (2007).

33 Rice, T. *et al.* Familial resemblance for body composition measures: The HERITAGE Family Study. *Obesity Research* **5**, 557-562 (1997).

34 Chaput, J.-P., Pérusse, L., Després, J.-P., Tremblay, A. & Bouchard, C. Findings from the Quebec Family Study on the Etiology of Obesity: Genetics and Environmental Highlights. *Current Obesity Reports* **3**, 54-66, doi:10.1007/s13679-013-0086-3 (2014).

35 Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457**, 480-U487, doi:10.1038/nature07540 (2009).

36 Menni, C. *et al.* Metabolomic markers reveal novel pathways of ageing and early development in human populations *International Journal of Epidemiology*, doi:10.1093/ije/dyt094 (2013).

37 Greenfield, J. R. *et al.* Oral glutamine increases circulating glucagon-like peptide 1, glucagon, and insulin concentrations in lean, obese, and type 2 diabetic subjects. *American Journal of Clinical Nutrition* **89**, 106-113, doi:10.3945/ajcn.2008.26362 (2009).

38 Pihlajamaki, J., Gylling, H., Miettinen, T. A. & Laakso, M. Insulin resistance is associated with increased cholesterol synthesis and decreased cholesterol absorption in normoglycemic men. *Journal of Lipid Research* **45**, 507-512, doi:10.1194/jlr.M300368-JLR200 (2004).

39 Gylling, H. *et al.* Insulin sensitivity regulates cholesterol metabolism to a greater extent than obesity: lessons from the METSIM Study. *Journal of Lipid Research* **51**, 2422-2427, doi:10.1194/jlr.P006619 (2010).

40 Grundberg, E. *et al.* Global Analysis of DNA Methylation Variation in Adipose Tissue from Twins Reveals Links to Disease-Associated Variants in Distal Regulatory Elements. *American Journal of Human Genetics* **93**, 876-890, doi:10.1016/j.ajhg.2013.10.004 (2013).

41 Montgomery, S. B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773-U151, doi:10.1038/nature08903 (2010).

42 Pickrell, J. K. *et al.* Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**, 768-772, doi:10.1038/nature08872 (2010).

43 Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology - Human gut microbes associated with obesity. *Nature* **444**, 1022-1023, doi:10.1038/nature4441022a (2006).

44 Ismail, N. A. *et al.* Frequency of Firmicutes and Bacteroidetes in gut microbiota in obese and normal weight Egyptian children and adults. *Archives of Medical Science* **7**, doi:10.5114/aoms.2011.23418 (2011).

45 Bervoets, L. *et al.* Differences in gut microbiota composition between obese and lean children: a cross-sectional study. *Gut Pathogens* **5**, doi:10.1186/1757-4749-5-10 (2013).

46 Payne, A. N. *et al.* The metabolic activity of gut microbiota in obese children is increased compared with normal-weight children and exhibits more exhaustive substrate utilization. *Nutrition & Diabetes* **1**, doi:10.1038/nutd.2011.8 (2011).

47 Finucane, M. M., Sharpton, T. J., Laurent, T. J. & Pollard, K. S. A Taxonomic Signature of Obesity in the Microbiome? Getting to the Guts of the Matter. *Plos One* **9**, doi:10.1371/journal.pone.0084689 (2014).

48 Turnbaugh, P. J. *et al.* The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice. *Science Translational Medicine* **1**, doi:10.1126/scitranslmed.3000322 (2009).

49 Ravussin, Y. *et al.* Responses of Gut Microbiota to Diet Composition and Weight Loss in Lean and Obese Mice. *Obesity* **20**, 738-747, doi:10.1038/oby.2011.111 (2012).

50 Clarke, S. F. *et al.* Exercise and associated dietary extremes impact on gut microbial diversity. *Gut* **63**, 1913-1920, doi:10.1136/gutjnl-2013-306541 (2014).

51 Tzatsos, A. & Kandror, K. V. Nutrients suppress phosphatidylinositol 3-kinase/Akt signaling via raptor-dependent mTOR-mediated insulin receptor substrate 1 phosphorylation. *Molecular and Cellular Biology* **26**, 63-76, doi:10.1128/mcb.26.1.63-76.2006 (2006).

52 Patti, M. E. & Kahn, B. B. Nutrient sensor links obesity with diabetes risk. *Nature Medicine* **10**, 1049-1050, doi:10.1038/nm1004-1049 (2004).

53 Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**, 335-336, doi:10.1038/nmeth.f.303 (2010).
54 Shin, S.-Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nature Genetics* **46**, 543-550, doi:10.1038/ng.2982 (2014).
55 Menni, C. *et al.* Metabolomic markers reveal novel pathways of ageing and early development in human populations *International Journal of Epidemiology*, In Press, doi:10.1093/ije/dyt094 (2013).
56 Boker, S. *et al.* OpenMx: An Open Source Extended Structural Equation Modeling Framework. *Psychometrika* **76**, 306-317, doi:10.1007/s11336-010-9200-6 (2011).
57 Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* **44**, 821-U136, doi:10.1038/ng.2310 (2012).
58 Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research* **42**, D1001-D1006, doi:10.1093/nar/gkt1229 (2014).
59 Teschendorff, A. E. *et al.* A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29**, 189-196, doi:10.1093/bioinformatics/bts680 (2013).
60 Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biology* **14**, doi:10.1186/gb-2013-14-10-r115 (2013).
61 Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353-1358, doi:10.1093/bioinformatics/bts163 (2012).

B.2 APPENDIX B2. GUT-MICROBIOTA-METABOLITE AXIS IN EARLY RENAL FUNCTION DECLINE

PLOS | ONE

RESEARCH ARTICLE

Gut-Microbiota-Metabolite Axis in Early Renal Function Decline

Clara Barrios[1,2☯]*, Michelle Beaumont[1☯], Tess Pallister[1], Judith Villar[3], Julia K. Goodrich[4], Andrew Clark[4], Julio Pascual[2], Ruth E. Ley[4], Tim D. Spector[1], Jordana T. Bell[1], Cristina Menni[1]*

1 Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom, **2** Department of Nephrology, Hospital del Mar, Institut Mar d'Investigacions Mediques, Barcelona, Spain, **3** Department of Infectious Diseases, Hospital del Mar, Institut Mar d'Investigacions Mediques, Barcelona, Spain, **4** Department of Molecular Biology and Genetics, Cornell University, Ithaca, United States of America

☯ These authors contributed equally to this work.
* CBarrios@parcdesalutmar.cat (CB); cristina.menni@kcl.ac.uk (CM)

OPEN ACCESS

Citation: Barrios C, Beaumont M, Pallister T, Villar J, Goodrich JK, Clark A, et al. (2015) Gut-Microbiota-Metabolite Axis in Early Renal Function Decline. PLoS ONE 10(8): e0134311. doi:10.1371/journal.pone.0134311

Editor: Giuseppe Remuzzi, Mario Negri Institute for Pharmacological Research and Azienda Ospedaliera Ospedali Riuniti di Bergamo, ITALY

Received: April 29, 2015

Accepted: July 7, 2015

Published: August 4, 2015

Data Availability Statement: All relevant data are within the paper.

Funding: TwinsUK was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013). The study also receives support from the National Institute for Health Research (NIHR) Clinical Research Facility at Guy's & St Thomas' NHS Foundation Trust and NIHR Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust, the King's College London, the Cornell Centre for Comparative Population Genomics. Tim Spector is the holder of an

Abstract

Introduction

Several circulating metabolites derived from bacterial protein fermentation have been found to be inversely associated with renal function but the timing and disease severity is unclear. The aim of this study is to explore the relationship between indoxyl-sulfate, p-cresyl-sulfate, phenylacetylglutamine and gut-microbial profiles in early renal function decline.

Results

Indoxyl-sulfate (Beta(SE) = -2.74(0.24); $P = 8.8x10^{-29}$), p-cresyl-sulfate (-1.99(0.24), $P = 4.6x10^{-16}$), and phenylacetylglutamine(-2.73 (0.25), $P = 1.2x10^{-25}$) were inversely associated with eGFR in a large population base cohort (TwinsUK, n = 4439) with minimal renal function decline. In a sub-sample of 855 individuals, we analysed metabolite associations with 16S gut microbiome profiles (909 profiles, QIIME 1.7.0). Three Operational Taxonomic Units (OTUs) were significantly associated with indoxyl-sulfate and 52 with phenylacetylglutamine after multiple testing; while one OTU was nominally associated with p-cresyl sulfate. All 56 microbial members belong to the order Clostridiales and are represented by anaerobic Gram-positive families Christensenellaceae, Ruminococcaceae and Lachnospiraceae. Within these, three microbes were also associated with eGFR.

Conclusions

Our data suggest that indoxyl-sulfate, p-cresyl-sulfate and phenylacetylglutamine are early markers of renal function decline. Changes in the intestinal flora associated with these metabolites are detectable in early kidney disease. Future efforts should dissect this relationship to improve early diagnostics and therapeutics strategies.

ERC Advanced Principal Investigator award. Clara Barrios is supported by a grant from the Spanish Society of Nephrology.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

It is increasingly recognized that the microbiome may affect health and disease of the host. Indeed the endogenous flora has been recently associated with type 2 diabetes, obesity, metabolic syndrome, cancer and liver cirrhosis among others [1–4]

Metabolites derived from bacteria provide a readout of the metabolic state of an individual and are the product of genetic [5,6] and exogenous (diet, lifestyle, gut microbial activity) factors under a particular set of conditions [7]. Under physiological conditions, there is a balance between the intestinal bacteria and the host, due to the innate immunity that maintains equilibrium in inflammation pathways and the intestinal barrier integrity. However, in chronic kidney disease (CKD), the uremic environment affects the intestinal barrier leading to bacterial dysbiosis [8]. This activates inflammatory pathways and immune processes and leads to systemic inflammation [9]. However, the degree of renal impairment that leads into modification of the intestinal milieu or the deficit of gut-metabolites excretion remains unclear.

A deeper understanding of the gut-microbe-metabolite axis is a prerequisite to improve therapeutic strategies that manipulate the gut microbiota in the onset of kidney dysfunction. Indoxyl-sulfate and p-cresyl-sulfate are end-products of bacterial protein fermentation of tryptophan and tyrosine respectively in the colon [10]. In vitro and ex vivo data show that indoxyl-sulfate and p-cresyl-sulfate may trigger or accelerate cardiovascular disease and progression of kidney failure [11,12]. Clinical observational studies also correlate high levels of both metabolites with overall mortality as well as cardiovascular disease and renal disease progression [13–15]. Phenylacetylglutamine is a major nitrogenous metabolite that accumulates in uremia. Its plasma levels increase after cigarette smoke exposure, in ischemic heart failure patients, hypertension, cardiovascular risk [16] and in the progression to end stage renal disease in type2 diabetic patients [17–19].

To date studies have concentrated on changes in intestinal flora and gut-metabolite levels in advanced stages of CKD [8,9,15,20–24], but potential changes in intestinal microbiota and gut microbial metabolites in early renal function decline have not yet been fully explored. To this end, we analyzed the links between metabolites indoxyl-sulfate, p-cresyl-sulfate and phenylacetylglutamine and gut microbiota to investigate whether changes at the individual operational taxonomic units (OTUs) level are detectable in early renal function decline.

Results and Discussion

Association of plasma circulating metabolites derived from bacterial protein fermentation was analyzed in 4439 individuals with different eGFR from the TwinsUK cohort. The demographic characteristics of the study populations are presented in **Table 1**. Out of 4439 individuals only 7.4% had eGFR<60 mL/min/1.73m^2. Indoxyl-sulfate (Beta(SE) = -2.74(0.24), $P = 8.8\text{x}10^{-29}$), p-cresyl-sulfate (-1.99(0.24), $P = 4.6\text{x}10^{-16}$), and phenylacetylglutamine (-2.73(0.25), $P = 1.2\text{x}10^{-25}$) were significantly and negatively associated with eGFR after adjusting for age, sex, body mass index (BMI), metabolite batch, type 2 diabetes, family relatedness and multiple testing using Bonferroni correction (**Table 2**).

As dietary factors are known to affect metabolites to varying levels [25,26], we tested their effect on the association between the metabolites and eGFR by including them as covariates in the linear model. Results were unchanged suggesting that dietary factors do not confound the three metabolite-eGFR association.

The plasma levels of these metabolites reflect the balance between elimination and generation. Some studies suggest most of the microbial derived metabolites are protein-bound [27], hence, elimination would depend on eGFR and the tubular transporter system.

Table 1. General Characteristics of the study population. Left column: Characteristics of population with renal and plasma metabolites data analyzed. Right column: Characteristics of sub-population with faecal microbiota data analyzed.

	Metabolites	Microbiota
Sample size, *n*	4439	855
Age, *yrs*	53.04±14.08	58.39±10.88
MZ:DZ:singletons	1795:1980:664	288:414:152
Female, *n (%)*	4162 (93.7)	840 (98.2)
BMI, *Kg/m²*	25.94±4.79	26.14±4.77
Creatinine, *mg/mL*	0.83±0.25	0.80±0.16
eGFR, *mL/min/1.73m²*	84.93±16.85	83.06±15.42
CKD (eGFR ≤ 60), *n (%)*	331 (7.4)	62 (7.2)
Type2 Diabetes, *n (%)*	78 (1.7)	21 (2.4)

CKD = Chronic Kidney Disease. eGFR = estimated glomerular filtration rate (CKD-EPI equation). MZ = monozygotic, DZ = dizygotic. Values for categorical variables are given as n (percentage); values for continuous variable as mean (± SD).

doi:10.1371/journal.pone.0134311.t001

A recent study, showed that eGFR provides an acceptable estimate of renal clearance of indoxyl and p-cresyl sulfate ($R^2 = 0.75$, $p<0.001$) in subjects with eGFR < 60mL/min/1.73m^2 [28]. These metabolites may be more sensitive to earlier stages of reduced renal function, as the eGFR-defined onset of CKD occurs only after half of the kidneys' filtration ability has been lost. Moreover, its higher levels in blood suggest the environmental changes affecting the intestinal flora could be playing a role in modifying the intestinal barrier before the onset of CKD.

We used 16S gut microbiome data available in a subset of the TwinsUK cohort individuals, to test for association between eGFR and plasma levels of indoxyl sulfate, p-cresyl sulfate and phenylacetylglutamine with 909 gut-microbial profiles (768 Operational Taxonomic Units (OTUs) and 141 collapsed taxonomies; see Methods). The gut microbiome 16s data have previously been described [29] and the current study analyzed a subset of 855 individuals with microbiome, fasting blood metabolites and eGFR data available (see demographic characteristics of the study population in Table 1). After adjusting for age, sex, BMI, metabolite batch, family relatedness and controlling for multiple testing using false discovery rate (FDR <5%), 3 OTUs were significantly associated with indoxyl-sulphate and 52 with phenylacetylglutamine (see Fig 1 and Table 3 for the full list). One OTU showed a borderline significance association with p-cresyl-sulphate but did not reach the FDR threshold. All the 56 microbial profiles belong to the order of Clostridiales and are mainly represented by the anaerobic Gram-positive families: Christensenellaceae, Ruminococcaceae and Lachnospiraceae. We then tested for association between these 56 microbies and renal function. After adjusting for covariates, 3 microbes were nominally associated with eGFR, and 2 were among those associated with phenylacetylglutamine and one with indoxyl-sulphate (see Fig 1 and Table 3 for

Table 2. Association and correlation of the metabolites and the eGFR.

Metabolites	eGFR		h^2[95%CI]*
	Beta(SE)	p	
Indoxyl-sulphate	-2.74 (0.24)	$8.8x10^{-29}$	0.24[0.12;0.37]
p-cresyl-sulphate	-1.99 (0.24)	$4.6x10^{-16}$	0.36[0.28;0.40]
Phenylacetylglutamine	-2.73 (0.25)	$1.2x10^{-25}$	0.33[0.21;0.44]

eGFR = estimated glomerular filtration rate. h^2 = *Heritability*.
*heritability estimates come from *Shin SY et al Nat Genet 2014 [6]*.

doi:10.1371/journal.pone.0134311.t002

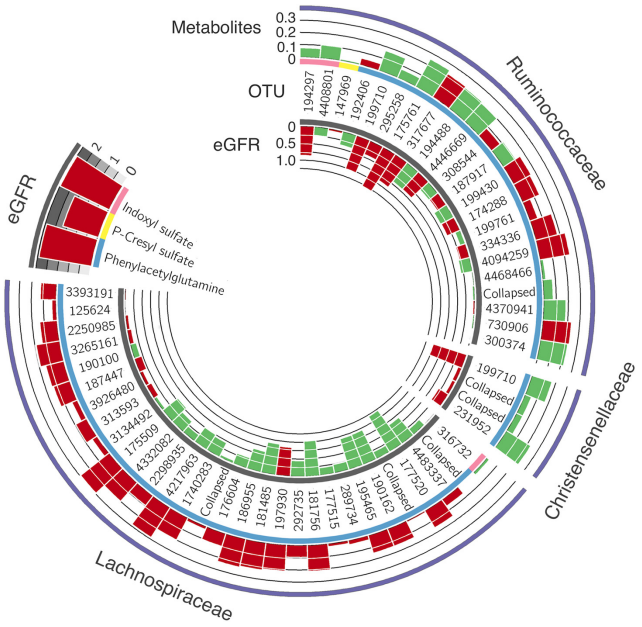

Fig 1. Circus histogram depicts positive and negative associations between the operational taxonomic units (OTUs) (middle circle), the metabolites (external circle) and the glomerular filtration rate (eGFR) (internal circle). Green color shows positive association while red color shows negative. The heights of the histogram columns represent the beta coefficients in the regression model. Upper left histograms represent the beta coefficients for the association between the three plasma metabolites and the eGFR. All the microbial traits belong to the order of Clostridiales and are represented by the families: Christensenellaceae, Ruminococcaceae and Lachnospiraceae.

doi:10.1371/journal.pone.0134311.g001

the full list). Microbes can also be affected by diet [29] and antibiotic use [30] and we therefore rerun the analyses adjusting for these confounders. Results were in line with those from the overall cohort, suggesting dietary pattern and antibiotic used are not affecting our associations. However, as data on diet and antibiotics was available for only 11% of the subjects with microbiota data, we cannot draw a more robust conclusion.

Previous studies showed that Ruminococcaceae, Lachnospiraceae and Christensenellaceae families are associated with healthier phenotypes. Indeed, Ruminococcaceae and Lachnospiraceae families have been found to be inversely associated with inflammatory bowel disease and are considered butyrate producers [31,32]. Butyrate is a preferred energy source for colonic epithelial cells and is thought to play an important role in maintaining colonic health in humans. Additionally, Christensenellaceae has been recently described by our group to be inversely correlated with BMI in humans and in experimental murine models [29]. In our data, a higher abundance of members of these three families was associated with lower circulating levels of indoxyl-sulphate, p-cresyl-sulphate and phenylacetylglutamine and related to better renal function. In line with our findings, a reduction in the number of culturable anaerobic bacteria

Table 3. Operational taxonomic units (OTUs) belonging to the order Clostridiales and represented by Ruminococcaceae, Christensenellaceae and Lachnospiraceae families. Table shows those significantly associated with Indoxyl sulfate and phenylacetylglutamine, the nominally associated with p-Cresyl sulfate and its association with the eGFR.

		Gut Microbes		h^2[95%CI]*	Metabolites			eGFR	
		Taxonomy			Indoxyl-Sulfate				
OTU	Order	Family	Genus.species		Beta(SE)	p	qval**	Beta(SE)	p
194297	Clostridiales	Ruminococcaceae	Ruminococcus	0.13 [0;0.25]	0.09(0.02)	0.0002	0.020	-0.86(0.44)	0.049
316732	Clostridiales	Lachnospiraceae	Lachnospira	0.2 [0;0.34]	0.04(0.01)	0.0006	0.036	0.02(0.20)	0.092
4408801	Clostridiales	Ruminococcaceae	Oscillospira	0.23 [0;0.35]	0.12(0.03)	0.0007	0.038	0.25(0.54)	0.063
					Metabolites			eGFR	
					p-Cresyl				
					Beta(SE)	p	qval**	Beta(SE)	p
147969	Clostridiales	Ruminococcaceae	Ruminococcus	0 [0;0.0]	0.03(0.009)	0.002	0.057	-0.06(0.22)	0.077
					Metabolites			eGFR	
					Phenylacetylglutamine				
					Beta(SE)	p	qval**	Beta(SE)	p
Collapsed	Clostridiales	Lachnospiraceae	Unclassified	0.21 [0;0.35]	-0.02(0.06)	0.0011	0.047	0.21(0.09)	0.028
4483337	Clostridiales	Lachnospiraceae	Unclassified	0 [0;0.3]	-0.08(0.02)	0.0002	0.021	0.67(0.33)	0.043
177520	Clostridiales	Lachnospiraceae	Roseburia faecis	0 [0;0.0]	-0.17(0.04)	0.0002	0.021	1.21(0.63)	0.056
192406	Clostridiales	Ruminococcaceae	Unclassified	0.12 [0;0.24]	-0.07(0.01)	$2.92x10^{-5}$	0.009	0.40(0.26)	0.133
Collapsed	Clostridiales	Lachnospiraceae	Unclassified	0.23 [0;0.35]	-0.03(0.01)	0.0007	0.036	0.22(0.15)	0.145
190162	Clostridiales	Lachnospiraceae	Blautia	0.16 [0;0.29]	-0.16(0.04)	0.0005	0.033	0.95(0.68)	0.163
199710	Clostridiales	Christensenellaceae	Unclassified	0.14 [0;0.29]	0.18(0.05)	0.0011	0.046	-1.03(0.76)	0.175
195465	Clostridiales	Lachnospiraceae	Unclassified	0.02 [0;0.25]	-0.16(0.05)	0.0009	0.043	0.94(0.71)	0.184
289734	Clostridiales	Lachnospiraceae	Unclassified	0.29 [0;0.41]	-0.05(0.01)	0.0002	0.019	0.25(0.19)	0.190
177515	Clostridiales	Lachnospiraceae	Roseburia	0 [0;0.24]	-0.04(0.01)	0.0009	0.041	0.22(0.17)	0.208
295258	Clostridiales	Ruminococcaceae	Unclassified	0.15 [0;0.28]	0.09(0.02)	0.0005	0.033	-0.43(0.35)	0.170
175761	Clostridiales	Ruminococcaceae	Unclassified	0.15 [0;0.29]	0.26(0.08)	0.0011	0.046	-1.3(1.09)	0.223
181756	Clostridiales	Lachnospiraceae	Blautia	0 [0;0.0]	-0.22(0.06)	0.0003	0.026	1.04(0.87)	0.232
292735	Clostridiales	Lachnospiraceae	Blautia	0.03 [0;0.36]	-0.11(0.02)	0.0002	0.019	0.46(0.40)	0.253
197930	Clostridiales	Lachnospiraceae	Unclassified	0.01 [0;0.14]	-0.22(0.05)	0.0001	0.015	-0.85(0.79)	0.283
317677	Clostridiales	Ruminococcaceae	Unclassified	0.06 [0;0.2]	-0.23(0.06)	0.0003	0.022	-0.91(0.93)	0.330
181485	Clostridiales	Lachnospiraceae	Ruminococcus	0 [0;0.0]	-0.23(0.07)	0.0010	0.044	0.91(0.95)	0.341
194488	Clostridiales	Ruminococcaceae	Unclassified	0 [0;0.0]	0.22(0.05)	0.0002	0.021	0.69(0.76)	0.365
4446669	Clostridiales	Ruminococcaceae	Unclassified	0 [0;0.0]	0.21(0.06)	0.0006	0.035	-0.68(0.82)	0.408
186955	Clostridiales	Lachnospiraceae	Unclassified	0.09 [0;0.23]	-0.24(0.05)	$2.69x10^{-5}$	0.009	0.62(0.78)	0.426
308544	Clostridiales	Ruminococcaceae	Unclassified	0.21 [0;0.38]	-0.10(0.03)	0.0011	0.047	0.33(0.43)	0.370
176604	Clostridiales	Lachnospiraceae	Unclassified	0.16 [0;0.29]	-0.12(0.03)	0.0002	0.019	0.32(0.43)	0.454
187917	Clostridiales	Ruminococcaceae	Unclassified	0.06 [0;0.26]	0.13(0.03)	0.0001	0.018	-0.31(0.46)	0.500
Collapsed	Clostridiales	Lachnospiraceae	Roseburia.uk	0.04 [0;0.63]	-0.04(0.01)	$9.33x10^{-6}$	0.006	0.08(0.13)	0.512
199430	Clostridiales	Ruminococcaceae	Faecalibacterium prausnitzii	0.12 [0;0.25]	-0.17(0.05)	0.0007	0.037	0.43(0.69)	0.530
174288	Clostridiales	Ruminococcaceae	Unclassified	0.03 [0;0.16]	-0.16(0.04)	0.0005	0.032	-0.41(0.66)	0.531
1740283	Clostridiales	Lachnospiraceae	Roseburia	0.04 [0;0.17]	-0.21(0.05)	0.0001	0.017	0.46(0.77)	0.545
Collapsed	Clostridiales	Christensenellaceae	Unclassified	0.38 [0.21;0.5]	0.07(0.01)	$3.54x10^{-7}$	0.001	-0.11(0.19)	0.550
4217963	Clostridiales	Lachnospiraceae	Unclassified	0.05 [0;0.18]	-0.29(0.06)	$1.26x10^{-5}$	0.006	0.51(0.9)	0.572
Collapsed	Clostridiales	Christensenellaceae	Unclassified	0.38 [0.21;0.49]	0.07(0.01)	$4.75x10^{-5}$	0.001	-0.10(0.19)	0.580
2298935	Clostridiales	Lachnospiraceae	Unclassified	0.01 [0;0.25]	-0.11(0.03)	0.0002	0.021	0.22(0.42)	0.604
4332082	Clostridiales	Lachnospiraceae	Roseburia	0.19 [0;0.31]	-0.21(0.05)	0.0003	0.026	0.40(0.79)	0.613
199761	Clostridiales	Ruminococcaceae	Unclassified	0.11 [0;0.24]	-0.08(0.02)	0.0006	0.033	0.16(0.32)	0.615
175509	Clostridiales	Lachnospiraceae	Blautia	0.01 [0;0.14]	-0.36(0.09)	0.0001	0.016	0.57(1.26)	0.647
3134492	Clostridiales	Lachnospiraceae	Unclassified	0.02 [0;0.24]	-0.06(0.01)	$3.64x10^{-5}$	0.016	0.09(0.20)	0.651

(*Continued*)

Table 3. (*Continued*)

231952	Clostridiales	Christensenellaceae	Unclassified	0.07 [0;0.21]	0.24(0.05)	$1.58x10^{-5}$	0.007	-0.33(0.76)	0.620
334336	Clostridiales	Ruminococcaceae	Unclassified	0.08 [0;0.21]	-0.19(0.05)	0.0009	0.042	-0.30(0.75)	0.668
313593	Clostridiales	Lachnospiraceae	Roseburia	0 [0;0.11]	-0.11(0.02)	0.0003	0.021	-0.16(0.40)	0.690
3926480	Clostridiales	Lachnospiraceae	Roseburia	0.06 [0;0.19]	-0.06(0.01)	0.0007	0.037	-0.10(0.25)	0.693
187447	Clostridiales	Lachnospiraceae	Roseburia	0.02 [0;0.15]	-0.15(0.04)	0.0005	0.032	-0.22(0.63)	0.723
4094259	Clostridiales	Ruminococcaceae	Unclassified	0.18 [0;0.3]	-0.25(0.07)	0.0008	0.039	0.34(1.05)	0.739
190100	Clostridiales	Lachnospiraceae	Blautia	0.15 [0;0.28]	-0.16(0.05)	0.0012	0.047	0.18(0.71)	0.797
4468466	Clostridiales	Ruminococcaceae	Unclassified	0.34 [0.1;0.45]	0.04(0.01)	0.0001	0.015	0.03(0.15)	0.809
3265161	Clostridiales	Lachnospiraceae	Unclassified	0.18 [0;0.31]	-0.20(0.05)	$4.89x10^{-5}$	0.012	-012(0.60)	0.848
4202174	Clostridiales	Clostridiaceae	Unclassified	0.15 [0;0.37]	0.07(0.02)	0.0008	0.040	-0.05(0.32)	0.864
Collapsed	Clostridiales	Ruminococcaceae	Oscillospira	0.13 [0;0.26]	0.09(0.02)	0.0006	0.034	0.06(0.37)	0.868
2250985	Clostridiales	Lachnospiraceae	Roseburia	0.10 [0;0.24]	-0.15(0.04)	0.0003	0.025	-0.08(0.55)	0.887
125624	Clostridiales	Lachnospiraceae	Unclassified	0.18 [0;0.15]	-0.05(0.01)	0.0001	0.015	0.02(0.21)	0.889
4370941	Clostridiales	Ruminococcaceae	Unclassified	0.28 [0.11;0.4]	0.19(0.04)	0.0001	0.015	-0.05(0.65)	0.929
3393191	Clostridiales	Lachnospiraceae	Roseburia	0.05 [0;0.18]	-0.14(0.04)	0.0007	0.038	-0.03(0.59)	0.957
730906	Clostridiales	Ruminococcaceae	Unclassified	0 [0;0.0]	-0.24(0.07)	0.0008	0.038	0.04(0.96)	0.960
300374	Clostridiales	Ruminococcaceae	Oscillospira	0.27 [0.11;0.3]	0.23(0.06)	0.0002	0.020	0.01(0.85)	0.983

* Heritability estimates comes from Goodrich JK et al. *Cell* 2014 [29].

**qval; is the significant threshold after apply false discovery rate (FDR <5%) adjustment. eGFR = estimated glomerular filtration rate.

doi:10.1371/journal.pone.0134311.t003

has been observed in CKD or on maintenance hemodialysis patients [33]. Our results suggest that CKD dysbiosis may start in earlier kidney function decline.

Heritability estimates for the three metabolites and the microbes identified are low/moderate heritability ranging from 0 to 0.38 (See Tables 2 and 3) suggesting that environmental factors have a major role in explaining the metabolite/microbe variation. Our heritability results are in line with those reported in non-twin population showing that metabolites derived from bacterial protein fermentation have low heritability [5].

Our study has some limitations. Firstly, the sample consists of predominantly healthy volunteer females with lower rate of diabetes and results may not be generalisable to males and to a population sample with greater prevalence of diabetes population. Moreover, estimates of GFR based on creatinine may underestimate renal function especially when GFR is >60 mL/min1.73m^2. Cystain C has been proposed as an alternative marker of renal function that could aid to reduce the bias. However, Cystatin C is not measured on the TwinsUK cohort. However, we have tried to minimize the underestimation bias using the CKD-EPI formula.

The cross-sectional nature of our data does not allow us to draw conclusions as to whether the findings are causative of kidney function decline or merely correlated with it. Finally, our study does not provide absolute quantification of the metabolites, and future studies are needed to establish reference ranges for clinical use.

To our knowledge, this is first study combining metabolome and microbiome data in early renal function decline. Our results have the potential to identify at risk patients before the onset of advanced CKD. Also, they open new avenues to our understanding of the renal-gut-microbiota-metabolite axis, which could improve therapeutic strategies. As well as providing early markers of renal damage, the microbiome can be manipulated allowing early therapeutic possibilities for prevention.

Concise Methods

Study subjects

Study subjects were twins enrolled in the Twins UK registry, a national register of adult twins started in 1992. The registry consists of over 10,000 predominantly female monozygotic and dizygotic twins, 18–84 years old, comparable to the general population in terms of lifestyle characteristics. Healthy twins were recruited from all over the UK as volunteers by successive media campaigns without selecting for particular diseases or traits. The TwinsUK cohort represents one of the most detailed omics and phenotypic resource in the word [34].

Data relevant to the present study include, BMI (body weight in kilograms divided by the square of height in square meters), type 2 Diabetes (t2D) (defined if fasting glucose $\geq$7 mmol/L or physician's letter confirming diagnosis). Renal parameters include estimated glomerular filtration rate (eGFR) calculated from standard creatinine using the CKD-EPI equation [35].

Dietary scores were obtain from food frequency questionnaires (FFQ) summarizing fruit and vegetable intake, alcohol intake, meat intake, hypo-caloric dieting and a "traditional English" diet as previously describe [25,26]. These five dietary scores are principal component analysis generated scores. As such they are independent variables standardized to have mean of zero and a SD of one in the whole TwinsUK study population. Each dietary pattern should be considered as the representative of a particular food pattern intake

Individuals were requested to complete a questionnaire regarding antibiotics used within the month previous faecal sample collection.

St. Thomas' Research Ethics Committee approved the study (EC96/439 TwinsUK) and all participants provided informed written consent.

Measurement of Metabolites

Non-targeted gas chromatography/mass spectrometry-based profiling was performed fasting plasma samples from participants in the TwinsUK cohort, using the Metabolon platform, as described previously [36,37]. Briefly, the Metabolon platform integrates the chemical analysis, including identification and relative quantification, data reduction, and quality assurance components of the process. This integrated platform enables the high-throughput collection and relative quantitative analysis of analytical data and identified a large number and broad spectrum molecules with a high degree of confidence. We inverse-normalised the metabolomics data and excluded metabolic traits with >20% missing values.

Microbiota analysis

Faecal samples were obtained from adult twin volunteers in the TwinsUK cohort. Faecal sample collection and 16S rRNA sequencing are described in depth previously in this sample (Goodrich et al) [29]. Briefly, the V4 region of the 16S rRNA gene was amplified and sequenced on Illumina MiSeq. Quality filtering and analysis of the sequence data with QIIME 1.7.0, was followed by closed-reference OTU picking to select OTUs at 97% sequence identity against the Greengenes May 2013 database as previously reported [38]. OTUs were adjusted for age, gender, shipment, number of sequences per sample and sequencing run. Collapsed taxonomic bins were created by combining OTUs of the same taxonomic designation into one variable. In total we used 768 OTUs and 141 collapsed taxonomies.

Statistical analysis

Statistical analysis was carried out using Stata version 12 and R version 3.1.2 (package LME4). Association analyses between eGFR and metabolites or microbiota profiles were performed

using random intercept linear regressions adjusting by age, sex, BMI, diabetes, experiment batch and family relatedness. Linear Mixed Effects Regression (LMER) was used to test the association between the microbiota and metabolites. Family structure and twin zygosity were accounted for as random effects and the microbe was the predictor variable. Multiple testing correction for the microbiota analysis was performed via false discovery rate (FDR<5%).

Author Contributions

Conceived and designed the experiments: CB MB JV CM. Performed the experiments: MB JG AC RL JB CM. Analyzed the data: CB MB TP CM. Contributed reagents/materials/analysis tools: MB JG AC JB CM. Wrote the paper: CB MB TP JV JP TS JB CM.

References

1. Han JL, Lin HL. Intestinal microbiota and type 2 diabetes: From mechanism insights to therapeutic perspective. World journal of gastroenterology: WJG, 20: 17737–17745, 2014.
2. Festi D, Schiumerini R, Eusebi LH, Marasco G, Taddia M, Colecchia A. Gut microbiota and metabolic syndrome. World journal of gastroenterology: WJG, 20: 16079–16094, 2014.
3. Monleon D, Morales JM, Barrasa A, Lopez JA, Vazquez C, Celda B. Metabolite profiling of fecal water extracts from human colorectal cancer. NMR in biomedicine, 22: 342–348, 2009.
4. Qin N, Yang F, Li A, Prifti E, Chen Y, Shao L, et al. Alterations of the human gut microbiome in liver cirrhosis. Nature, 513: 59–64, 2014.
5. Viaene L, Thijs L, Jin Y, Liu Y, Gu Y, Meijers B, et al. Heritability and clinical determinants of serum indoxyl sulfate and p-cresyl sulfate, candidate biomarkers of the human microbiome enterotype. PloS one, 9: e79682, 2014.
6. Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang J, et al. An atlas of genetic influences on human blood metabolites. Nature genetics, 46: 543–550, 2014.
7. Holmes E, Wilson ID, Nicholson JK. Metabolic phenotyping in health and disease. Cell, 134: 714–717, 2008.
8. Anders HJ, Andersen K, Stecher B. The intestinal microbiota, a leaky gut, and abnormal immunity in kidney disease. Kidney international, 83: 1010–1016, 2013.
9. Vaziri ND. CKD impairs barrier function and alters microbial flora of the intestine: a major link to inflammation and uremic toxicity. Current opinion in nephrology and hypertension, 21: 587–592, 2012.
10. Meyer TW, Hostetter TH. Uremic solutes from colon microbes. Kidney international, 81: 949–954, 2012.
11. Niwa T. Indoxyl sulfate is a nephro-vascular toxin. Journal of renal nutrition: the official journal of the Council on Renal Nutrition of the National Kidney Foundation, 20: S2–6, 2010.
12. Tumur Z, Shimizu H, Enomoto A, Miyazaki H, Niwa T. Indoxyl sulfate upregulates expression of ICAM-1 and MCP-1 by oxidative stress-induced NF-kappaB activation. American journal of nephrology, 31: 435–441, 2010.
13. Poesen R, Viaene L, Verbeke K, Augustijns P, Bammens B, Claes K, et al. Cardiovascular disease relates to intestinal uptake of p-cresol in patients with chronic kidney disease. BMC nephrology, 15: 87, 2014.
14. Barreto FC, Barreto DV, Liabeuf S, Meert N, Glorieux G, Temmar M, et al. European Uremic Toxin Work, G: Serum indoxyl sulfate is associated with vascular disease and mortality in chronic kidney disease patients. Clinical journal of the American Society of Nephrology: CJASN, 4: 1551–1558, 2009.
15. Wu IW, Hsu KH, Lee CC, Sun CY, Hsu HJ, Tsai CJ, et al. p-Cresyl sulphate and indoxyl sulphate predict progression of chronic kidney disease. *Nephrology*, *dialysis*, *transplantation: official publication of the* European Dialysis and Transplant Association—European Renal Association, 26: 938–947, 2011.
16. Menni C, Mangino M, Cecelja M, Psatha M, Brosnan MJ, Trimmer J, et al. Metabolomic study of carotid-femoral pulse-wave velocity in women. Journal of hypertension, 33: 791–796, 2015.
17. Cruickshank-Quinn CI, Mahaffey S, Justice MJ, Hughes G, Armstrong M, Bowler RP, et al. Transient and persistent metabolomic changes in plasma following chronic cigarette smoke exposure in a mouse model. PloS one, 9: e101855, 2014.
18. Kang SM, Park JC, Shin MJ, Lee H, Oh J, Ryu do H, et al. (1)H nuclear magnetic resonance based metabolic urinary profiling of patients with ischemic heart failure. Clinical biochemistry, 44: 293–299, 2011.

19. Niewczas MA, Sirich TL, Mathew AV, Skupien J, Mohney RP, Warram JH, et al. Uremic solutes and risk of end-stage renal disease in type 2 diabetes: metabolomic study. Kidney international, 85: 1214–1224, 2014.

20. Meijers BK, Claes K, Bammens B, de Loor H, Viaene L, Verbeke K, et al. p-Cresol and cardiovascular risk in mild-to-moderate kidney disease. Clinical journal of the American Society of Nephrology: CJASN, 5: 1182–1189, 2010.

21. Lin CJ, Chen HH, Pan CF, Chuang CK, Wang TJ, Sun FJ, et al. p-Cresylsulfate and indoxyl sulfate level at different stages of chronic kidney disease. Journal of clinical laboratory analysis, 25: 191–197, 2011.

22. Ramezani A, Raj DS. The gut microbiome, kidney disease, and targeted interventions. Journal of the American Society of Nephrology: JASN, 25: 657–670, 2014.

23. Aronov PA, Luo FJ, Plummer NS, Quan Z, Holmes S, Hostetter TH, et al. Colonic contribution to uremic solutes. Journal of the American Society of Nephrology: JASN, 22: 1769–1776, 2011.

24. Vaziri ND, Wong J, Pahl M, Piceno YM, Yuan J, DeSantis TZ, et al. Chronic kidney disease alters intestinal microbial flora. Kidney international, 83: 308–315, 2013.

25. Teucher B, Skinner J, Skidmore PM, Cassidy A, Fairweather-Tait SJ, Hooper L, et al. Dietary patterns and heritability of food choice in a UK female twin cohort. Twin research and human genetics: the official journal of the International Society for Twin Studies, 10: 734–748, 2007.

26. Menni C, Zhai G, Macgregor A, Prehn C, Romisch-Margl W, Suhre K, et al. Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. Metabolomics: Official journal of the Metabolomic Society, 9: 506–514, 2013.

27. Meijers BK, Bammens B, Verbeke K, Evenepoel P, A review of albumin binding in CKD. American journal of kidney diseases: the official journal of the National Kidney Foundation, 51: 839–850, 2008.

28. Poesen R, Viaene L, Verbeke K, Claes K, Bammens B, Sprangers B, et al. Renal clearance and intestinal generation of p-cresyl sulfate and indoxyl sulfate in CKD. Clinical journal of the American Society of Nephrology: CJASN, 8: 1508–1514, 2013.

29. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, et al. Human genetics shape the gut microbiome. Cell, 159: 789–799, 2014.

30. Angelakis E, Merhej V, Raoult D. Related actions of probiotics and antibiotics on gut microbiota and weight modification. The Lancet Infectious diseases, 13: 889–899, 2013.

31. Vital M, Howe AC, Tiedje JM. Revealing the bacterial butyrate synthesis pathways by analyzing (meta) genomic data. mBio, 5: e00889, 2014.

32. Kostic AD, Xavier RJ, Gevers D. The microbiome in inflammatory bowel disease: current status and the future ahead. Gastroenterology, 146: 1489–1499, 2014.

33. Ranganathan N, Friedman EA, Tam P, Rao V, Ranganathan P, Dheer R. Probiotic dietary supplementation in patients with stage 3 and 4 chronic kidney disease: a 6-month pilot scale trial in Canada. Current medical research and opinion, 25: 1919–1930, 2009.

34. Moayyeri A, Hammond CJ, Valdes AM, Spector TD. Cohort Profile: TwinsUK and healthy ageing twin study. International journal of epidemiology, 42: 76–85, 2013.

35. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, Feldman HI, et al. Ckd, EPI: A new equation to estimate glomerular filtration rate. Annals of internal medicine, 150: 604–612, 2009.

36. Menni C, Kastenmuller G, Petersen AK, Bell JT, Psatha M, Tsai PC, et al. Metabolomic markers reveal novel pathways of ageing and early development in human populations. International journal of epidemiology, 42: 1111–1119, 2013.

37. Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. Analytical chemistry, 81: 6656–6667, 2009.

38. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nature methods, 7: 335–336, 2010.

B.3 APPENDIX B3. HUMAN GENETICS SHAPE THE GUT MICROBIOME.

Article

Cell

Human Genetics Shape the Gut Microbiome

Julia K. Goodrich,[1,2] Jillian L. Waters,[1,2] Angela C. Poole,[1,2] Jessica L. Sutter,[1,2] Omry Koren,[1,2,7] Ran Blekhman,[1,8] Michelle Beaumont,[3] William Van Treuren,[4] Rob Knight,[4,5,6] Jordana T. Bell,[3] Timothy D. Spector,[3] Andrew G. Clark,[1] and Ruth E. Ley[1,2,*]

[1]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA
[2]Department of Microbiology, Cornell University, Ithaca, NY 14853, USA
[3]Department of Twin Research and Genetic Epidemiology, King's College London, London SE1 7EH, UK
[4]Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA
[5]Biofrontiers Institute, University of Colorado, Boulder, CO 80309, USA
[6]Howard Hughes Medical Institute, University of Colorado, Boulder, CO 80309, USA
[7]Present address: Faculty of Medicine, Bar Ilan University, Safed 13115, Israel
[8]Present address: Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, MN 55455, USA
*Correspondence: rel222@cornell.edu
http://dx.doi.org/10.1016/j.cell.2014.09.053

SUMMARY

Host genetics and the gut microbiome can both influence metabolic phenotypes. However, whether host genetic variation shapes the gut microbiome and interacts with it to affect host phenotype is unclear. Here, we compared microbiotas across >1,000 fecal samples obtained from the TwinsUK population, including 416 twin pairs. We identified many microbial taxa whose abundances were influenced by host genetics. The most heritable taxon, the family Christensenellaceae, formed a co-occurrence network with other heritable Bacteria and with methanogenic Archaea. Furthermore, Christensenellaceae and its partners were enriched in individuals with low body mass index (BMI). An obese-associated microbiome was amended with *Christensenella minuta*, a cultured member of the Christensenellaceae, and transplanted to germ-free mice. *C. minuta* amendment reduced weight gain and altered the microbiome of recipient mice. Our findings indicate that host genetics influence the composition of the human gut microbiome and can do so in ways that impact host metabolism.

INTRODUCTION

The human gut microbiome has been linked to metabolic disease and obesity (Karlsson et al., 2013; Le Chatelier et al., 2013; Ley et al., 2005; Qin et al., 2012; Turnbaugh et al., 2009). Variation in host genetics can also underlie susceptibility to metabolic disease (Frayling et al., 2007; Frazer et al., 2009; Herbert et al., 2006; Yang et al., 2012). Despite these shared effects, the relationship between host genetic variation and the diversity of gut microbiomes is largely unknown.

The gut microbiome is environmentally acquired from birth (Costello et al., 2012; Walter and Ley, 2011), therefore it may function as an environmental factor that interacts with host genetics to shape phenotype, as well as a genetically determined attribute that is shaped by, and interacts with, the host (Bevins and Salzman, 2011; Spor et al., 2011; Tims et al., 2011). Because the microbiome can be modified for therapeutic applications (Borody and Khoruts, 2012; Hamilton et al., 2013; Khoruts et al., 2010; van Nood et al., 2013), it constitutes an attractive target for manipulation. Once the interactions between host genetics and the microbiome are understood, its manipulation could be optimized for a given host genome to reduce disease risk.

Although gut microbiomes can differ markedly in diversity across adults (Human Microbiome Project Consortium, 2012; Qin et al., 2010), family members are often observed to have more similar microbiotas than unrelated individuals (Lee et al., 2011; Tims et al., 2013; Turnbaugh et al., 2009; Yatsunenko et al., 2012). Familial similarities are usually attributed to shared environmental influences, such as dietary preference, a powerful shaper of microbiome composition (Cotillard et al., 2013; David et al., 2014; Wu et al., 2011). Yet related individuals share a larger degree of genetic identity, raising the possibility that shared genetic composition underlies familial microbiome similarities.

Support for a host genetic effect on the microbiome comes mostly from studies taking a targeted approach. For instance, the concordance rate for carriage of the methanogen *Methanobrevibacter smithii* is higher for monozygotic (MZ) than dizygotic (DZ) twin pairs (Hansen et al., 2011), and studies comparing microbiotas between human subjects differing at specific genetic loci have shown gene-microbiota interactions (Frank et al., 2011; Khachatryan et al., 2008; Rausch et al., 2011; Rehman et al., 2011; Wacklin et al., 2011). A more general approach to this question has linked genetic loci with abundances of gut bacteria in mice (Benson et al., 2010; McKnite et al., 2012), but in humans, a general approach (e.g., using twins) has failed to reveal significant genotype effects on microbiome diversity (Turnbaugh et al., 2009; Yatsunenko et al., 2012). Thus, heritable components of the human gut microbiome remain to be identified using an unbiased approach.

Here, we assessed the heritability of the gut microbiome with a well-powered twin study. Comparisons between MZ and DZ twin

Cell

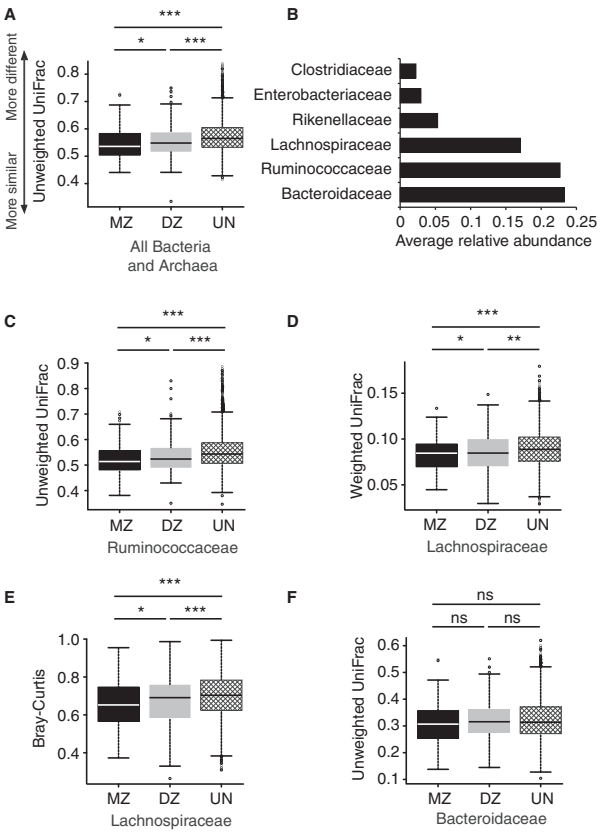

Figure 1. Microbiomes Are More Similar for Monozygotic Than Dizygotic Twins

(A and C–F) Boxplots of β diversity distances between microbial communities obtained when comparing individuals within twinships for monozygotic (MZ) twin pairs and dizygotic (DZ) twin pairs, and between unrelated individuals (UN). (A) The whole microbiome. (C) The bacterial family Ruminococcaceae. (D and E) The bacterial family Lachnospiraceae. (F) The family Bacteroidaceae. The specific distance metric used in each analysis is indicated on the axes. *p < 0.05, **p < 0.01, ***p < 0.001 for Student's t tests with 1,000 Monte Carlo simulations.

(B) The average relative abundances in the whole data set for the top six most prevalent bacterial families (unrarefied data, see Experimental Procedures).

See also Figure S1 and Table S1.

pairs allowed us to assess the impact of genotype and early shared environment on their gut microbiota. Our study addressed the following questions: Which specific taxa within the gut microbiome are heritable, and to what extent? Which predicted metagenomic functions are heritable? How do heritable microbes relate to host BMI? Finally, we use fecal transplants into germ-free mice to assess the phenotype effects of the most heritable taxon.

RESULTS

Twin Data Set

We obtained 1,081 fecal samples from 977 individuals: 171 MZ and 245 DZ twin pairs, two from twin pairs with unknown zygosity, and 143 samples from just one twin within a twinship (i.e., unrelated). In addition, we collected longitudinal samples from 98 of these individuals (see Supplemental Information available online). Most subjects were female, ranging in age from 23 to 86 years (average age: 60.6 ± 0.3 years). The average BMI of the subjects was 26.25 (± 0.16) with the following distribution: 433 subjects had a low to normal BMI (<25), 322 had an overweight BMI (25-30), 183 were obese (>30), and 39 individuals in which the current BMI status was unknown. We generated 78,938,079 quality-filtered sequences that mapped to the Bacteria and Archaea in the Greengenes database (average sequences per sample: 73,023 ± 889).

Microbiome Composition and Richness

We sorted sequences into 9,646 operational taxonomic units (OTUs, ≥97% ID). Of these OTUs, 768 were present in at least 50% of the samples. Taxonomic classification revealed a fairly typical Western diversity profile: the dominant bacterial phyla were Firmicutes (53.9% of total sequences), Bacteroidetes (35.3%), Proteobacteria (4.5%), with Verrucomicrobia, Actinobacteria, and Tenericutes each comprising 2% of the sequences, and a tail of rare bacterial phyla that together accounted for the remaining 1% of the sequences.

The most widely shared methanogen was *M. smithii* (64% of people, using nonrarefied data), followed by vadinCA11, a member of the Thermoplasmata with no cultured representatives (~6%), *Methanosphaera stadtmanae* (~4%), and *Methanomassiliicoccus* (~4%, a member of the Thermoplasmata). Forty-six of the 61 samples in which we detected vadinCA11 also contained *M. smithii*, indicating that the two most dominant archaeal taxa are not mutually exclusive. Faith's PD was positively correlated with the relative abundance of the family Methanobacteriaceae (rho = 0.42 rarefied, 0.37 for transformed

Cell

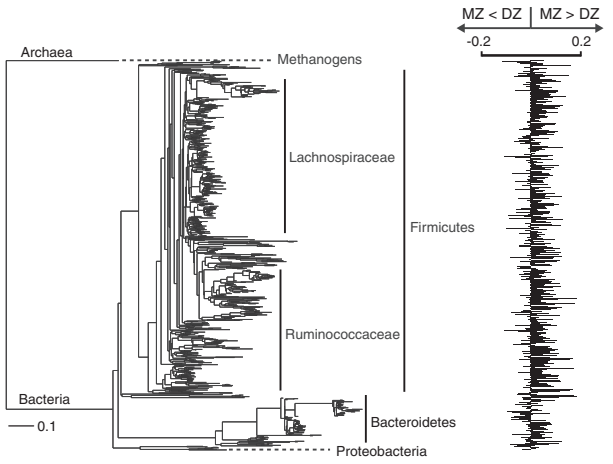

Figure 2. OTU Relative Abundances Are More Highly Correlated within MZ Than DZ Twin Pairs

Left: a phylogeny of taxa in the TwinsUK study (Greengenes tree pruned to include only OTUs shared by 50% of the TwinsUK participants). Right: corresponding twin-pair intraclass correlation coefficients (ICCs). ICCs were calculated for each OTU and the difference in correlation coefficients for MZ twin pairs versus DZ twin pairs. Bars pointing to the right indicate that the difference is positive (i.e., MZ ICCs > DZ ICCs) and bars pointing to the left indicate negative differences (DZ ICCs > MZ ICCs). The scale bar associated with the phylogeny shows substitutions/site. See also Figure S2.

counts, $p < 1 \times 10^{-11}$), which corroborates previous reports of higher richness associating with methanogens.

Broad Diversity Comparisons between MZ and DZ Twin Pairs

We observed that microbiotas were more similar overall within individuals (resampled) than between unrelated individuals ($p < 0.001$ for weighted and unweighted UniFrac and Bray-Curtis using a Student's t test with 1,000 Monte Carlo simulations) (Table S1A) and were also more similar within twin pairs compared to unrelated individuals ($p < 0.009$ for weighted and unweighted UniFrac and Bray-Curtis) (Figures 1 and S1; Table S1). MZ twin pairs had more similar microbiotas than DZ twins for the unweighted UniFrac metric ($p = 0.032$), but not the weighted UniFrac and Bray-Curtis metrics (Figures 1A and S1). As greater similarities for MZ versus DZ twin pairs are seen in unweighted UniFrac but not abundance-based metrics, MZ similarities are driven by shared community membership rather than structure.

We next constrained the distance metric analyses to the three most dominant bacterial families: the Lachnospiraceae and Ruminococcaceae (Firmicutes) and Bacteroidaceae (Figure 1B). We observed greater similarities for MZ compared to DZ twins using the unweighted UniFrac metric within the Ruminococcaceae family (Figure 1C). Within the Lachnospiraceae family, significantly greater similarity for MZ compared to DZ twins emerged using the weighted UniFrac and Bray-Curtis metrics (Figures 1D and 1E). In contrast, when restricted to the Bacteroidaceae family, we found that MZ and DZ twins had similar pairwise diversity using all three metrics (Figures 1F, S1B, and S1E).

MZ Twins Have More Highly Correlated Microbiotas

We next asked if the abundances of specific taxa were generally more highly correlated within MZ twin pairs compared to DZ twin pairs. For each twin pair we generated intraclass correlation coefficients (ICCs) for the relative abundances of OTUs. Mean ICCs were significantly greater for MZ compared to DZ twin pairs (Wilcoxon signed rank test on ICCs at the OTU level, $p = 6 \times 10^{-04}$; Figure 2). Because many OTUs are closely phylogenetically related, their abundances may not be independent, which may inflate levels of significance. To account for this effect, we maintained the structure of the phylogenetic tree but permuted the MZ and DZ labels in 10,000 tests to generate randomized ICCs. As an independent validation, we also applied these analyses to two previously published data sets generated originating in a population of twins from Missouri, USA: "Turnbaugh" (Turnbaugh et al., 2009), which described 54 twin pairs ranging from 21 to 32 years of age, and "Yatsunenko" (Yatsunenko et al., 2012), which included 63 twin pairs with an age range of 13–30 years of age. Mean ICCs of OTU abundances were significantly greater for MZ compared to DZ twin pairs in both of these data sets (significance by permutation: $p < 0.001$ and 0.047 respectively; Figure S2), corroborating our observations.

Heritability Estimates for OTUs and Predicted Functions

We estimated heritability using the twin-based ACE model, which partitions the total variance into three component sources: genetic effects (A), common environment (C), and unique environment (E) (Eaves et al., 1978). The largest proportion of variance in abundances of OTUs could be attributed to the twins' unique environments (i.e., E > A; Table S2). However, for the majority of OTUs (63%), the proportion of variance attributed to genetic effects was greater than the proportion of variance attributed to common environment (A > C; Table S2).

From the ACE model, we calculated 95% confidence intervals for the heritability estimates and determined the significance of the heritability values using a permutation method to generate nominal p values (Table S2). We found a high correlation between the tail probability for inclusion of zero in the confidence interval of heritability and the p values obtained from the permutation tests (rho = 0.872, $p < 10^{-15}$), indicating substantial consistency across these tests. Although heritability studies

Cell

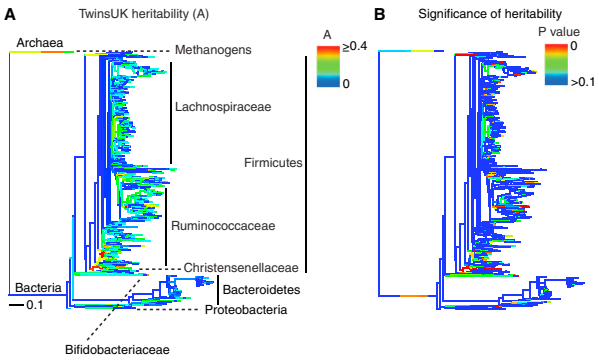

Figure 3. Heritability of Microbiota in the TwinsUK Data Set

(A) OTU Heritability (A from ACE model) estimates mapped onto a microbial phylogeny and displayed using a rainbow gradient from blue (A = 0) to red (A ≥ 0.4). This phylogenetic tree was obtained from the Greengenes database and pruned to include only nodes for which at least 50% of the TwinsUK participants were represented.

(B) The significance for the heritability values shown in (A) was determined using a permutation test (n = 1,000) and are shown on the same phylogeny as in (A). P values range from 0 (red) to >0.1 (blue).

See also Figure S3 and Table S2.

traditionally report confidence intervals and nominal p values only, we also generated FDR-corrected p values (Table S2).

We also applied the ACE model to the abundances of sequences mapping to each node in the phylogeny. Across the three studies, the nodes of the phylogeny with the strongest heritabilities lie within the Ruminococcaceae and Lachnospiraceae families, and the Bacteroidetes are mostly environmentally determined (Figures 3 and S3). Subsets of the Archaea are also heritable in the TwinsUK and the Yatsunenko studies (the Turnbaugh study did not include data for Archaea).

We characterized the longitudinal stability of each OTU by calculating the ICCs of the OTU abundance across repeat samples, which consisted of two samples collected from the same individual at different times. By permuting these repeat sample ICCs, we found that heritable OTUs (A > 0.2) were more stable (ICC > 0.6) than expected by chance (Figure S3E; $p < 0.001$, p value was determined as the fraction of permutations that had greater than or equal to the observed number of OTUs that are both heritable and stable).

We used PICRUSt (Langille et al., 2013) to produce predicted metagenomes from the 16S rRNA gene sequence data and applied the ACE model to estimate the heritability of predicted abundances of conserved orthologous groups (COGs). This analysis revealed six functions with heritabilities A > 0.2 and nominal p values < 0.05 (p values are generated by permutation testing; Extended Experimental Procedures; Table S2). Correcting for multiple comparisons, one category, "secondary metabolites biosynthesis, transport and catabolism" (Q), passed a stringent FDR (A = 0.32, 95% confidence interval [CI] = 0.16–0.44). We also tested α diversity for heritability and found that it was not heritable.

The Family Christensenellaceae Is the Most Highly Heritable Taxon

The most heritable taxon overall was the family Christensenellaceae (A = 0.39, 95% CI = 0.21–0.49, p = 0.001; Figure 4A; Table S2; this taxon passes a stringent FDR) of the order Clostridiales. Christensenellaceae was also highly heritable in the Yatsunenko data set (A = 0.62, 95% CI = 0.38–0.77; Figure 4B; Table S2). We repeated this analysis for the taxa abundances with the effect of BMI regressed out, and results were highly correlated (Pearson correlation = 0.95, $p < 1 \times 10^{-15}$).

Christensenellaceae Is the Hub in a Co-Occurrence Network with Other Heritable Taxa

We observe a module of co-occurring heritable families, and the hub (node connected to most other nodes) of this module is the family Christensenellaceae (Figures 5A and S4A). The heritable module includes the families Methanobacteriaceae (Archaea) and Dehalobacteriaceae (Firmicutes) and the orders SHA-98 (Firmicutes), RF39 (Tenericutes), and ML615J-28 (Tenericutes). The Christensenellaceae network is anticorrelated with the Bacteroidaceae and Bifidobacteriaceae families. We validated these results by applying this method to the family-level taxonomic abundances in the Yatsunenko data set (as this one is most technically similar to the TwinsUK data set), where we also found the same Christensenellaceae-centered module of heritable families anticorrelated to the Bacteroidaceae/Bifidobacteriaceae module (Figure S4B).

Christensenellaceae Associates with a Low BMI

The family Christensenellaceae was significantly enriched in subjects with a lean BMI (<25) compared to those with an obese BMI (>30; Benjamini-Hochberg corrected p value < 0.05 from t test on transformed counts; Table S2). Other members of the Christensenellaceae consortium were also enriched in lean-BMI subjects: the Dehalobacteriaceae, SHA-98, RF39, and the Methanobacteriaceae (Figure 5B). Overall, a majority (n = 35) of the OTUs with highest heritability scores (A > 0.2, nominal p < 0.05) were enriched in the lean subjects. A subset of OTUs classified as *Oscillospira* were enriched in lean subjects, and *M. smithii*, although not significantly heritable, was positively associated with a lean BMI.

Christensenellaceae Is Associated with Health in Published Data Sets

Because the names *Christensenella* and Christensenellaceae were only recently assigned to the bacterial phylogeny, we

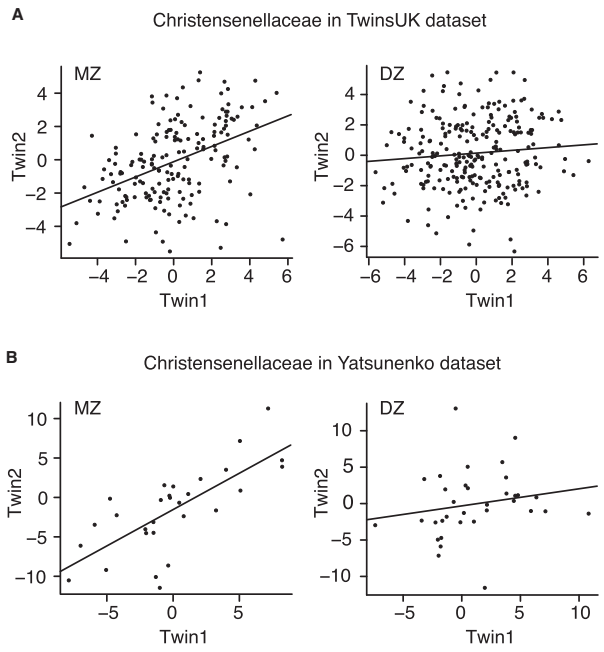

Figure 4. MZ Twin Pairs Have Higher Correlations of Christensenellaceae Than DZ Twin Pairs in TwinsUK and Yatsunenko Data Sets

Scatter plots comparing the abundances of Christensenellaceae in the gut microbiota of MZ and DZ co-twins. Christensenellaceae abundances were transformed and adjusted to control for technical and other covariates (Residuals are plotted, see Extended Experimental Procedures) and the data are separated by zygosity (MZ or DZ twins).

(A) TwinsUK data set.

(B) Yatsunenko data set.

assessed the abundances of sequences assigned to these taxa in previously published studies. This analysis revealed that members of the Christensenellaceae were enriched in fecal samples of healthy versus pediatric and young adult IBD patients ($p < 0.05$) (Papa et al., 2012). Christensenellaceae were at greater abundance in lean BMI compared to obese-BMI twins in the Turnbaugh data set, but the difference was not quite significant ("time-point 2" samples, $p = 0.07$). In a case study of the development of an infant's gut microbiome (Koenig et al., 2011), Christensenellaceae was present at 8.6% in the mother's stool at the time of birth and at 20% in the infant's meconium. We also noted that Christensenellaceae is enriched in omnivorous compared to herbivorous and carnivorous mammals (Muegge et al., 2011). However, we did not find a relationship between Christensenellaceae and diet information in human studies (Wu et al., 2011; Martínez et al., 2010; Koren et al., 2012).

Christensenellaceae Is Associated with Reduced Weight Gain in Germ-free Mice Inoculated with Lean and Obese Human Fecal Samples

Methanogens co-occurred with Christensenellaceae in this study and have been linked to low BMI in previous studies. Because of this previous association with a low-BMI, we wanted to ensure that methanogens were present in the Christensenellaceae consortium in an initial experiment exploring its effect on weight phenotypes. Therefore, we selected 21 donors for fecal transfer to germ-free mice based on BMI status (low or high) and presence or absence of the methanogen-Christensenellaceae consortium. Donors fell into one of four categories: lean with detectable methanogens (L+), lean without methanogens (L−), obese with methanogens (O+), or obese without methanogens (O−). The abundance of Christensenellaceae positively correlated with the abundance of methanogens in donor stool (rho = 0.72, $p = 0.0002$), indicating that methanogen abundance was a good proxy for the methanogen-Christensenellaceae consortium.

A 16S rRNA analysis of the fecal microbiomes before and after transfer to germ-free mice showed that although members of the Christensenellaceae were present throughout the experiment in recipient mice (Figure 6A), *M. smithii* was undetectable in the mouse fecal or cecal samples (the first sampling was at 20 hr postinoculation). At 20 hr postinoculation, the microbiota had shifted dramatically in diversity from the inoculation, but by day 5 had shifted back partially and remained fairly stable through day 21 (Figures 6B, 6C, S5A, and S5B). Abundances of *Christensenella* were correlated with PC3 (abundances rarefied at 55,000 sequences per sample versus unweighted UniFrac; Spearman rho = 0.59, $p < 2.2 \times 10^{-16}$), and PC3 captured the differences between the four donor groups (Figure 6D). We observed a trend for *Christensenella* abundances as highest in the L+ group and lowest in the O− group (Figure 6A), which mirrored the weight differences between those groups: the percent change in body weights of the recipient mice was significantly lower in the L+ group compared to the O− group (day 12, $p < 0.05$, t test; Figures 6E and 6F). Cecal levels of propionate and butyrate were significantly elevated in mice receiving methanogen-positive compared to methanogen-negative microbiomes controlling for the effect of donor BMI (two-way ANOVA, $p < 0.05$ for both SCFAs; Figures S5C–S5E). Stool energy content was significantly higher for the methanogen-positive microbiomes at day 12, when the percent changes in weight were greatest (two-way ANOVA, $p = 0.004$,

Cell

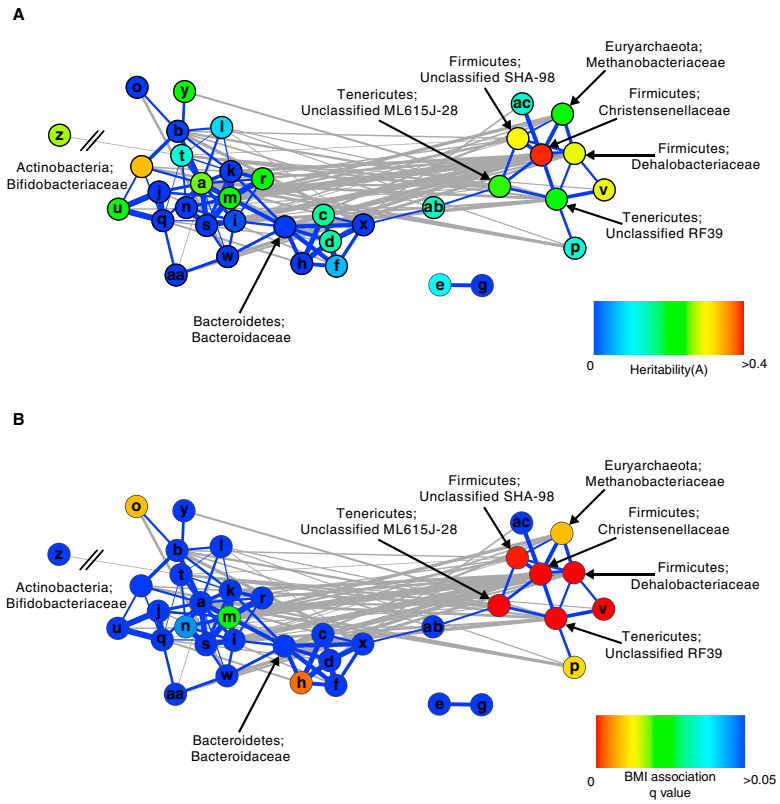

Figure 5. Christensenellaceae Is the Hub of a Consortium of Co-occurring Heritable Microbes that Are Associated with a Lean BMI

The same network built from SparCC correlation coefficients between sequence abundances collapsed at the family level. The nodes represent families and the edges represent the correlation coefficients between families. Edges are colored blue for a positive correlation and gray for a negative correlation, and the weight of the edge reflects the strength of the correlation. Nodes are positioned using an edge-weighted force directed layout.

(A) Nodes are colored by the heritability of the family.

(B) Nodes are colored by the significance of the association families and a normal versus obese BMI. Family names are either indicated on the panel, or nodes are given a letter code. Phylum Actinobacteria: (a) Actinomycetaceae, (b) Coriobacteriaceae; Phylum Bacteroidetes: (c) Barnesiellaceae, (d) Odoribacteraceae, (e) Paraprevotellaceae, (f) Porphyromonadaceae, (g) Prevotellaceae, (h) Rikenellaceae; Phylum Firmicutes: (i) Carnobacteriaceae, (j) Clostridiaceae, (k) Erysipelotrichaceae, (l) Eubacteriaceae, (m) Lachnospiraceae, (n) Lactobacillaceae, (o) Mogibacteriaceae, (p) Peptococcaceae, (q) Peptostreptococcaceae, (r) Ruminococcaceae, (s) Streptococcaceae, (t) Tissierellaceae, (u) Turicibacteraceae, (v) Unclassified Clostridiales, (w) Veillonellaceae; Phylum Proteobacteria: (x) Alcaligenaceae, (y) Enterobacteriaceae, (z) Oxalobacteraceae, (aa) Pasteurellaceae, (ab) Unclassified RF32; Phylum Verrucomicrobia: (ac) Verrucomicrobiaceae. See also Figure S4.

no effect of BMI or interaction; Figure S5F). In a replicated experiment, using 21 new donors, the same weight differences were observed (a significantly lower mean weight gain for the L+ compared to the O− mouse recipients at day 10 postinoculation; one-way t test, p = 0.047; Figure S5G).

Christensenella minuta Added to Donor Stool Reduces Adiposity Gains in Recipient Mice

Based on the observation that *Christensenella* levels in the previous experiment were similar to the weight gain patterns, we performed experiments in which a donor stool lacking

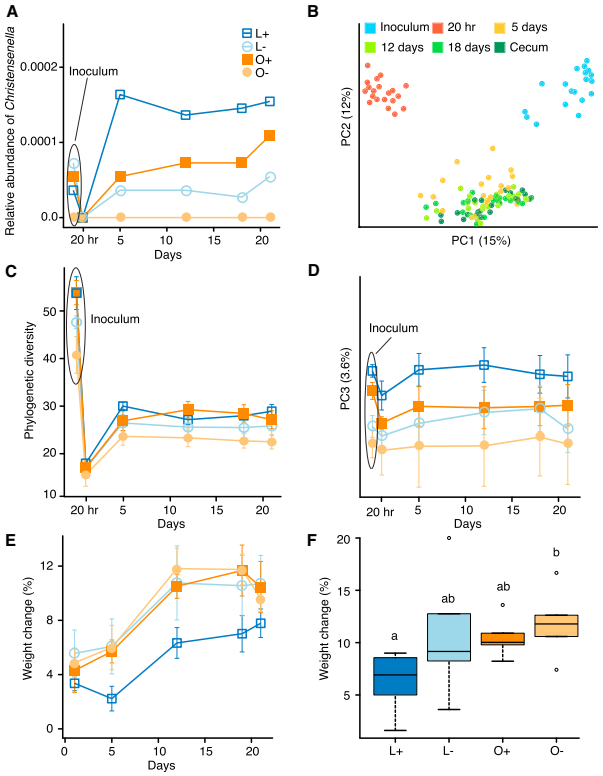

Figure 6. Fecal Transplants from Obese and Lean UK Twins to Germ-Free Mice Reveal Levels of Christensenellaceae Posttransfer Mirror Delayed Weight Gain

(A) Median relative abundances for OTUs classified as the genus *Christensenella* in the four donor treatment groups over time in the recipient mouse microbiotas.

(B) Principal coordinates analysis of unweighted UniFrac distances for (1) the inoculum prior to transplantation, (2) fecal samples at four time points, and (3) cecal samples at day 21 posttransplant; see panel legend for color key. The amount of variance described by the first two PCs is shown on the axes.

(C) Mean values ± SEM for richness (Faith's PD) for the microbiomes of the transplant mice plotted against time (days postinoculation, with day 0 = inoculation day).

(D) The mean values ± SEM for PC3 derived for the same analysis as shown in (B) are plotted against time (day 0 = inoculation day) for the four treatment groups. The amount of variance explained by PC3 is in parentheses.

(E) Percent weight change since inoculation for germ-free mouse recipients of 21 donor stools that were obtained from lean or obese donors with or without detectable *M. smithii*, which was used as a marker for the Christensenellaceae consortium. Means for each treatment group are plotted ± SEM.

(F) Boxplots for percent weight changes for the four groups at day 12 posttransplant, when maximal weight differences were observed. Letters next to boxes indicate significant differences if letters are different ($p < 0.05$). For all panels: dark blue, L+, lean donor with methanogens; light blue, L−, lean donor lacking methanogens; dark orange, O+, obese donor with methanogens; light orange, O−, obese donor without methanogens. We repeated this experiment with a set of 21 new mice and unique human donors and recovered the same effect.

See also Figure S5.

detectable *Christensenella* was amended with *C. minuta* and weight gain of recipient mice was monitored. One obese human donor was selected from the 21 donors from the first transplant experiment based on its lack of detectable OTUs assigned to the genus *Christensenella*. At day 21 postgavage, mice receiving the *C. minuta* treatment weighed significantly less than those that received unamended stool (nested ANOVA, $p < 0.05$; Figure 7A). Adiposity was significantly lower for mice receiving the *C. minuta* treatment (nested ANOVA, $p = 9.4 \times 10^{-5}$, Figure 7B). Energy content for stool collected at day 21 was not different between treatments (data not shown).

Analysis of the microbial community by 16S rRNA gene sequencing showed an impact on the overall community diversity that persisted over time (Figures 7C and 7D). After an initial acclimation (20 hr), the communities within recipient mice began to separate by treatment regardless of the effects of time and cocaging (Figures 7C, 7D, and S6). At 5 days postinoculation, the relative abundance of *C. minuta* was similar to that observed in the previous transplant experiment and persisted throughout the duration of the study. We identified two genera that discriminated the two treatments at day 21: *Oscillospira* and a genus within the Rikenellaceae were enriched in the *C. minuta* treatment (misclassification error rate of 0.06). *Oscillospira* abundances were significantly correlated with PC2 in the unweighted UniFrac analysis of the communities (rho = −0.71, $p = 0.0009$; Figure 7E), which is the PC that separates the *C. minuta*-amended and unamended microbiotas.

DISCUSSION

Our results represent strong evidence that the abundances of specific members of the gut microbiota are influenced in part by the genetic makeup of the host. Earlier studies using fingerprinting approaches also reported host genetic effects (Stewart

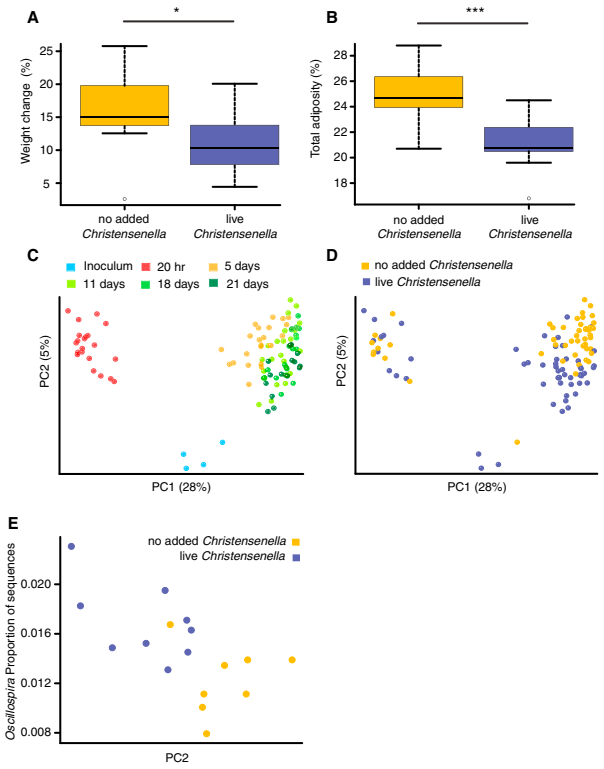

Figure 7. Addition of *Christensenella minuta* to Donor Stool Leads to Reduced Weight and Adiposity Gains in Recipient Mice

(A) Boxplot of percent weight change for germ-free mouse recipients of a single donor stool only (lacking detectable *Christensenella* in unrarefied 16S rRNA data) or the donor stool amended with live *C. minuta*.

(B) Boxplots showing percent body fat for mice in each group at day 21 (n = 12 mice per treatment).

(C and D) Principal coordinates analysis of unweighted UniFrac distances for (1) the inoculum prior to transplantation, (2) fecal samples at five time points posttransplant; see legend for color key. The amount of variance described by the first two PCs is shown on the axes. The same data projection is shown in (C) and (D); sample symbols are colored by time point (C) and by treatment (D).

(E) Relationship between PCs from the PCoA analysis and levels of *Oscillospira* at day 21 (rho = −0.71, p = p < 0.001). Symbols are colored by treatment.

See also Figure S6.

et al., 2005; Zoetendal et al., 2001), but without sequence data it is not possible to know if the taxa shown here to be heritable were also driving those patterns. The Turnbaugh et al. (2009) and Yatsunenko et al. (2012) studies, which are quite similar in experimental approach, reported a lack of host genetic effect on the gut microbiome, most likely because both studies were underpowered. Nevertheless, reanalysis of the data from both studies validated our observation that the abundances of taxa are more highly correlated within MZ than DZ twin pairs. Thus, host genetic interactions with specific taxa are likely widespread across human populations, with profound implications for human biology.

The most highly heritable taxon in our data set was the family Christensenellaceae, which was also the hub of a co-occurrence network that includes other taxa with high heritability. A notable component of this network was the archaeal family Methanobacteriaceae. Similarly, Hansen et al. (2011) had previously identified members of the Christensenellaceae (reported as relatives of *Catabacter*) as co-occurring with *M. smithii*. These co-occurrence patterns could derive from different scenarios: for instance, multiple taxa may be heritable and co-occur while each taxon is affected by host genetics independently, or alternatively one (or a few) taxa may be heritable and other taxa correlate with host genetics due to their co-occurrence with these key heritable taxa. Further experimental research will be required to elucidate if the co-occurring heritable taxa interact in syntrophic partnerships or simply respond similarly to host-influenced environmental cues in the gut.

Our results suggest that environmental factors mostly shape the Bacteroidetes community, because most were not heritable. These results are consistent with those of a recent study of Finnish MZ twins, in which levels of *Bacteroides* spp. were more similar between twins when their diets were similar (Simões et al., 2013). Members of the Bacteroidetes have been shown to respond to diet interventions (Wu et al., 2011; David et al., 2014)

Importantly, the family Christensenellaceae is heritable in the Yatsunenko data set and its network is also present. This validation did not involve a directed search using the taxa identified in this study but was made by applying the ACE model independently. In the TwinsUK as well as the Missouri twins data sets, the majority of OTUs with the highest heritability estimates fell within the Ruminococcaceae and Lachnospiraceae families. The Missouri and TwinsUK studies differed somewhat in the levels and structure of heritability, which may relate to study size (Kuczynski et al., 2010), participant age (Claesson et al., 2011), population (Yatsunenko et al., 2012), and/or diet (Wu et al., 2011), all of which have been shown to affect microbiome structure.

Cell

The high heritability of the Christensenellaceae raises questions about the nature of interactions between the host and members of this family, but to date there is little published work with which to infer their roles. *Christensenella minuta* is Gram-negative, nonspore forming, nonmotile, and produces SCFAs (Morotomi et al., 2012). A review of publicly available data suggests it is present from birth and associates with a healthy state but not with diet. Thus, although diet is a heritable trait in the same population (Menni et al., 2013; Teucher et al., 2007), it does not appear to be driving the heritability of the Christensenellaceae. Obesity is also strongly heritable in the TwinsUK population, raising the question of whether the heritabilities we saw for gut microbes were driven by BMI. To test this, we reran the heritability calculations using residuals after regressing out the effect of BMI and found that results of the two analyses were highly correlated. This suggests that the effect of host genetics on Christensenellaceae abundance is independent of an effect of BMI.

Our transplantation experiments showed a moderating effect of methanogen-presence in the human donor on weight gain of recipient mice, although strikingly, *M. smithii* did not persist in mice. In contrast, Christensenellaceae levels in mice mirrored their weight gain. Transfer to germ-free mice of microbiomes from obese and lean donors generally results in greater adiposity gains for obese compared to lean donors (Ridaura et al., 2013; Turnbaugh et al., 2008; Vijay-Kumar et al., 2010). These studies have not reported the methanogen or Christensenellaceae status of the donors, so whether these microbes affect the host phenotype is unknown. *M. smithii* has been associated with a lean phenotype in multiple studies (Million et al., 2012, 2013; Schwiertz et al., 2010; Armougom et al., 2009; Le Chatelier et al., 2013), raising the possibility that methanogens are key components of the consortium for regulating host phenotype. The results of our methanogen-Christensenellaceae transfer revealed that although methanogens may be a marker for a low BMI in humans, they are not required to promote a lean phenotype in the germ-free mouse experimental model. This result suggests that methanogens may be functionally replaced by another consortium member in the mouse, while the Christensenellaceae are not.

The results of the *C. minuta* spike-in experiments supported the hypothesis that members of the Christensenellaceae promote a lean host phenotype. Addition of *C. minuta* also remodeled the diversity of the community. Intriguingly, *Oscillospira*, which includes heritable OTUs in the TwinsUK data set and is associated with a lean BMI, was enriched in the *C. minuta*-amended microbiomes. How *C. minuta* reshapes the community remains to be explored. The relatively low levels of *C. minuta* and its profound effects on the community and the host may indicate that it is a keystone taxon. Together these findings indicate that the Christensenellaceae are highly heritable bacteria that can directly contribute to the host phenotype with which they associate.

Conclusions

Host genetic variation drives phenotype variation, and this study solidifies the notion that our microbial phenotype is also influenced by our genetic state. We have shown that the host genetic effect varies across taxa and includes members of different phyla. The host alleles underlying the heritability of gut microbes, once identified, should allow us to understand the nature of our association with these health-associated bacteria and eventually to exploit them to promote health.

EXPERIMENTAL PROCEDURES

Human Subjects and Sample Collection

Fecal samples were obtained from adult twin pair participants of the TwinsUK registry (Moayyeri et al., 2013). Most participants were women (only 20 men were included). Twins collected fecal samples at home, and the samples were refrigerated for up to 2 days prior to their annual clinical visit at King's College London, at which pointed they were stored at −80°C until processing.

Diversity and Phylogenetic Analyses

We amplified 16S rRNA genes (V4) from bulk DNA by PCR prior to sequencing on the Illumina MiSeq 2 × 250 bp platform at Cornell Biotechnology Resource Center Genomics Facility. We performed quality filtering and analysis of the 16S rRNA gene sequence data with QIIME 1.7.0 (Caporaso et al., 2010).

Predicted Metagenomes

PICRUSt v1.0.0 was used to predict abundances of COGs from the OTU abundances rarefied at 10,000 sequences per sample.

Heritability Estimations

Heritability estimates were calculated on the OTU abundances, taxon bins, nodes throughout the bacterial phylogenetic tree, α-diversity, and PICRUSt-predicted COGs using the structural equation modeling software OpenMx (Boker et al., 2011).

Microbiota Transfer Experiments

Stool samples from the Twins UK cohort were selected as described in the main text and inoculated into 6-week-old germ-free Swiss Webster mice via oral gavage, with one recipient mouse per fecal donor. Mice were single-housed. For the *Christensenella minuta* addition, three experiments were conducted: In the first experiment, one treatment group received only donor stool inoculum, whereas the other received donor stool amended with 1×10^8 *C. minuta* cells; for the second experiment, a heat-killed *C. minuta* control was added; in the third experiment, the heat-killed control was compared to live *C. minuta* only (no vehicle-only control). Mice were housed four per cage, with three cages per treatment. In all experiments, mice were provided with autoclaved food and water ad libitum and maintained on a 12 hr light/dark cycle. Body weight and chow consumption were monitored and fecal pellets were collected weekly. At sacrifice, adiposity was analyzed using DEXA. Body, mesenteric adipose tissue, and gonadal fat pad tissue weights were collected at this time. Gross energy content of mouse stool samples was measured by bomb calorimetry using an IKA C2000 calorimeter (Dairy One). Wet cecal contents were weighed and resuspended in 2% (v/v) formic acid by vortexing and measured using a gas chromatograph (HP series 6890) with flame ionization detection.

Statistical Analysis

Values are expressed as the mean ± SEM unless otherwise indicated. Full methods are described in the Supplemental Information.

ACCESSION NUMBERS

The European Bioinformatics Institute (EBI) accession numbers for the sequences reported in this paper are ERP006339 and ERP006342.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2014.09.053.

Cell

AUTHOR CONTRIBUTIONS

R.E.L. and A.G.C. supervised the study. J.T.B. and T.D.S. helped design study and provided comments and discussion. J.T.B. and T.D.S. oversaw collection of samples. J.K.G., R.E.L., O.K., J.L.S., A.C.P., and J.L.W. oversaw microbial data generation. J.K.G. performed the analysis with contributions from R.E.L., R.B., A.G.C., J.L.W., O.K., A.C.P., M.B., W.V.T., and R.K. J.K.G. and J.L.W. performed microbiota transfer experiments. J.K.G., J.L.W., and R.E.L. prepared the manuscript with comments from A.G.C., T.D.S., J.T.B., R.B., and R.K.

ACKNOWLEDGMENTS

We thank Wei Zhang, Sara Di Rienzi, Lauren Harroff, Largus Angenent, Hannah de Jong, Gabe Fox, Nick Scalfone, Aymé Spor, and Beth Bell for their help. We also thank three anonymous reviewers for their helpful comments and Mary-Claire King for her advice and encouragement. This work was funded by NIH RO1 DK093595, DP2 OD007444, The Cornell Center for Comparative Population Genomics, the Wellcome Trust, and the European Community's Seventh Framework Programme (FP7/2007-2013). The study also received support from the National Institute for Health Research (NIHR) BioResource Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. R.E.L. is a Fellow of the David and Lucile Packard Foundation and of the Arnold and Mabel Beckman Foundation. J.K.G. is a National Academy of Sciences predoctoral Fellow. T.D.S. is holder of an ERC Advanced Researcher Award. R.K. is a Howard Hughes Medical Institute Early Career Scientist.

Received: April 3, 2014
Revised: July 10, 2014
Accepted: September 24, 2014
Published: November 6, 2014

REFERENCES

Armougom, F., Henry, M., Vialettes, B., Raccah, D., and Raoult, D. (2009). Monitoring bacterial community of human gut microbiota reveals an increase in Lactobacillus in obese patients and Methanogens in anorexic patients. PLoS ONE *4*, e7125.

Benson, A.K., Kelly, S.A., Legge, R., Ma, F., Low, S.J., Kim, J., Zhang, M., Oh, P.L., Nehrenberg, D., Hua, K., et al. (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. Proc. Natl. Acad. Sci. USA *107*, 18933–18938.

Bevins, C.L., and Salzman, N.H. (2011). The potter's wheel: the host's role in sculpting its microbiota. Cell. Mol. Life Sci. *68*, 3675–3685.

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., Spies, J., Estabrook, R., Kenny, S., Bates, T., et al. (2011). OpenMx: An open source extended structural equation modeling framework. Psychometrika *76*, 306–317.

Borody, T.J., and Khoruts, A. (2012). Fecal microbiota transplantation and emerging applications. Nat. Rev. Gastroenterol. Hepatol. *9*, 88–96.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods *7*, 335–336.

Claesson, M.J., Cusack, S., O'Sullivan, O., Greene-Diniz, R., de Weerd, H., Flannery, E., Marchesi, J.R., Falush, D., Dinan, T., Fitzgerald, G., et al. (2011). Composition, variability, and temporal stability of the intestinal microbiota of the elderly. Proc. Natl. Acad. Sci. USA *108* (*Suppl 1*), 4586–4591.

Costello, E.K., Stagaman, K., Dethlefsen, L., Bohannan, B.J., and Relman, D.A. (2012). The application of ecological theory toward an understanding of the human microbiome. Science *336*, 1255–1262.

Cotillard, A., Kennedy, S.P., Kong, L.C., Prifti, E., Pons, N., Le Chatelier, E., Almeida, M., Quinquis, B., Levenez, F., Galleron, N., et al.; ANR MicroObes consortium (2013). Dietary intervention impact on gut microbial gene richness. Nature *500*, 585–588.

David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A., et al. (2014). Diet rapidly and reproducibly alters the human gut microbiome. Nature *505*, 559–563.

Eaves, L.J., Last, K.A., Young, P.A., and Martin, N.G. (1978). Model-fitting approaches to the analysis of human behaviour. Heredity (Edinb) *41*, 249–320.

Frank, D.N., Robertson, C.E., Hamm, C.M., Kpadeh, Z., Zhang, T., Chen, H., Zhu, W., Sartor, R.B., Boedeker, E.C., Harpaz, N., et al. (2011). Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. Inflamm. Bowel Dis. *17*, 179–184.

Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., Perry, J.R., Elliott, K.S., Lango, H., Rayner, N.W., et al. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science *316*, 889–894.

Frazer, K.A., Murray, S.S., Schork, N.J., and Topol, E.J. (2009). Human genetic variation and its contribution to complex traits. Nat. Rev. Genet. *10*, 241–251.

Hamilton, M.J., Weingarden, A.R., Unno, T., Khoruts, A., and Sadowsky, M.J. (2013). High-throughput DNA sequence analysis reveals stable engraftment of gut microbiota following transplantation of previously frozen fecal bacteria. Gut Microbes *4*, 125–135.

Hansen, E.E., Lozupone, C.A., Rey, F.E., Wu, M., Guruge, J.L., Narra, A., Goodfellow, J., Zaneveld, J.R., McDonald, D.T., Goodrich, J.A., et al. (2011). Pan-genome of the dominant human gut-associated archaeon, Methanobrevibacter smithii, studied in twins. Proc. Natl. Acad. Sci. USA *108* (*Suppl 1*), 4599–4606.

Herbert, A., Gerry, N.P., McQueen, M.B., Heid, I.M., Pfeufer, A., Illig, T., Wichmann, H.E., Meitinger, T., Hunter, D., Hu, F.B., et al. (2006). A common genetic variant is associated with adult and childhood obesity. Science *312*, 279–283.

Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207–214.

Karlsson, F.H., Tremaroli, V., Nookaew, I., Bergström, G., Behre, C.J., Fagerberg, B., Nielsen, J., and Bäckhed, F. (2013). Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature *498*, 99–103.

Khachatryan, Z.A., Ktsoyan, Z.A., Manukyan, G.P., Kelly, D., Ghazaryan, K.A., and Aminov, R.I. (2008). Predominant role of host genetics in controlling the composition of gut microbiota. PLoS ONE *3*, e3064.

Khoruts, A., Dicksved, J., Jansson, J.K., and Sadowsky, M.J. (2010). Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent Clostridium difficile-associated diarrhea. J. Clin. Gastroenterol. *44*, 354–360.

Koenig, J.E., Spor, A., Scalfone, N., Fricker, A.D., Stombaugh, J., Knight, R., Angenent, L.T., and Ley, R.E. (2011). Succession of microbial consortia in the developing infant gut microbiome. Proc. Natl. Acad. Sci. USA *108* (*Suppl 1*), 4578–4585.

Koren, O., Goodrich, J.K., Cullender, T.C., Spor, A., Laitinen, K., Bäckhed, H.K., Gonzalez, A., Werner, J.J., Angenent, L.T., Knight, R., et al. (2012). Host remodeling of the gut microbiome and metabolic changes during pregnancy. Cell *150*, 470–480.

Kuczynski, J., Costello, E.K., Nemergut, D.R., Zaneveld, J., Lauber, C.L., Knights, D., Koren, O., Fierer, N., Kelley, S.T., Ley, R.E., et al. (2010). Direct sequencing of the human microbiome readily reveals community differences. Genome Biol. *11*, 210.

Langille, M.G., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Vega Thurber, R.L., Knight, R., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat. Biotechnol. *31*, 814–821.

Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J.-M., Kennedy, S., et al.; MetaHIT Consortium (2013). Richness of human gut microbiome correlates with metabolic markers. Nature *500*, 541–546.

Lee, S., Sung, J., Lee, J., and Ko, G. (2011). Comparison of the gut microbiotas of healthy adult twins living in South Korea and the United States. Appl. Environ. Microbiol. *77*, 7433–7437.

Ley, R.E., Bäckhed, F., Turnbaugh, P., Lozupone, C.A., Knight, R.D., and Gordon, J.I. (2005). Obesity alters gut microbial ecology. Proc. Natl. Acad. Sci. USA *102*, 11070–11075.

Lozupone, C.A., Hamady, M., Kelley, S.T., and Knight, R. (2007). Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. Appl. Environ. Microbiol. *73*, 1576–1585.

Martínez, I., Kim, J., Duffy, P.R., Schlegel, V.L., and Walter, J. (2010). Resistant starches types 2 and 4 have differential effects on the composition of the fecal microbiota in human subjects. PLoS ONE *5*, e15046.

McKnite, A.M., Perez-Munoz, M.E., Lu, L., Williams, E.G., Brewer, S., Andreux, P.A., Bastiaansen, J.W., Wang, X., Kachman, S.D., Auwerx, J., et al. (2012). Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. PLoS ONE *7*, e39191.

Menni, C., Zhai, G., Macgregor, A., Prehn, C., Römisch-Margl, W., Suhre, K., Adamski, J., Cassidy, A., Illig, T., Spector, T.D., and Valdes, A.M. (2013). Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. Metabolomics *9*, 506–514.

Million, M., Maraninchi, M., Henry, M., Armougom, F., Richet, H., Carrieri, P., Valero, R., Raccah, D., Vialettes, B., and Raoult, D. (2012). Obesity-associated gut microbiota is enriched in Lactobacillus reuteri and depleted in Bifidobacterium animalis and Methanobrevibacter smithii. Int J Obes (Lond) *36*, 817–825.

Million, M., Angelakis, E., Maraninchi, M., Henry, M., Giorgi, R., Valero, R., Vialettes, B., and Raoult, D. (2013). Correlation between body mass index and gut concentrations of Lactobacillus reuteri, Bifidobacterium animalis, Methanobrevibacter smithii and Escherichia coli. Int J Obes (Lond) *37*, 1460–1466.

Moayyeri, A., Hammond, C.J., Valdes, A.M., and Spector, T.D. (2013). Cohort Profile: TwinsUK and healthy ageing twin study. Int J oEpidemiol *42*, 76–85.

Morotomi, M., Nagai, F., and Watanabe, Y. (2012). Description of Christensenella minuta gen. nov., sp. nov., isolated from human faeces, which forms a distinct branch in the order Clostridiales, and proposal of Christensenellaceae fam. nov. Int. J. Syst. Evol. Microbiol. *62*, 144–149.

Muegge, B.D., Kuczynski, J., Knights, D., Clemente, J.C., González, A., Fontana, L., Henrissat, B., Knight, R., and Gordon, J.I. (2011). Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science *332*, 970–974.

Papa, E., Docktor, M., Smillie, C., Weber, S., Preheim, S.P., Gevers, D., Giannoukos, G., Ciulla, D., Tabbaa, D., Ingram, J., et al. (2012). Non-invasive mapping of the gastrointestinal microbiota identifies children with inflammatory bowel disease. PLoS ONE *7*, e39242.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., et al.; MetaHIT Consortium (2010). A human gut microbial gene catalogue established by metagenomic sequencing. Nature *464*, 59–65.

Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature *490*, 55–60.

Rausch, P., Rehman, A., Künzel, S., Häsler, R., Ott, S.J., Schreiber, S., Rosenstiel, P., Franke, A., and Baines, J.F. (2011). Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. Proc. Natl. Acad. Sci. USA *108*, 19030–19035.

Rehman, A., Sina, C., Gavrilova, O., Häsler, R., Ott, S., Baines, J.F., Schreiber, S., and Rosenstiel, P. (2011). Nod2 is essential for temporal development of intestinal microbial communities. Gut *60*, 1354–1362.

Ridaura, V.K., Faith, J.J., Rey, F.E., Cheng, J., Duncan, A.E., Kau, A.L., Griffin, N.W., Lombard, V., Henrissat, B., Bain, J.R., et al. (2013). Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science *341*, 1241214.

Schwiertz, A., Taras, D., Schäfer, K., Beijer, S., Bos, N.A., Donus, C., and Hardt, P.D. (2010). Microbiota and SCFA in lean and overweight healthy subjects. Obesity (Silver Spring) *18*, 190–195.

Simões, C.D., Maukonen, J., Kaprio, J., Rissanen, A., Pietiläinen, K.H., and Saarela, M. (2013). Habitual dietary intake is associated with stool microbiota composition in monozygotic twins. J. Nutr. *143*, 417–423.

Spor, A., Koren, O., and Ley, R. (2011). Unravelling the effects of the environment and host genotype on the gut microbiome. Nat. Rev. Microbiol. *9*, 279–290.

Stewart, J.A., Chadwick, V.S., and Murray, A. (2005). Investigations into the influence of host genetics on the predominant eubacteria in the faecal microflora of children. J. Med. Microbiol. *54*, 1239–1242.

Teucher, B., Skinner, J., Skidmore, P.M., Cassidy, A., Fairweather-Tait, S.J., Hooper, L., Roe, M.A., Foxall, R., Oyston, S.L., Cherkas, L.F., et al. (2007). Dietary patterns and heritability of food choice in a UK female twin cohort. Twin Res. Hum. Genet. *10*, 734–748.

Tims, S., Zoetendal, E.G., Vos, W.M., and Kleerebezem, M. (2011). Host genotype and the effect on microbial communities. In Metagenomics of the Human Body, K.E. Nelson, ed. (New York: Springer), pp. 15–41.

Tims, S., Derom, C., Jonkers, D.M., Vlietinck, R., Saris, W.H., Kleerebezem, M., de Vos, W.M., and Zoetendal, E.G. (2013). Microbiota conservation and BMI signatures in adult monozygotic twins. ISME J. *7*, 707–717.

Turnbaugh, P.J., Bäckhed, F., Fulton, L., and Gordon, J.I. (2008). Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. Cell Host Microbe *3*, 213–223.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., Sogin, M.L., Jones, W.J., Roe, B.A., Affourtit, J.P., et al. (2009). A core gut microbiome in obese and lean twins. Nature *457*, 480–484.

van Nood, E., Vrieze, A., Nieuwdorp, M., Fuentes, S., Zoetendal, E.G., de Vos, W.M., Visser, C.E., Kuijper, E.J., Bartelsman, J.F., Tijssen, J.G., et al. (2013). Duodenal infusion of donor feces for recurrent Clostridium difficile. N. Engl. J. Med. *368*, 407–415.

Vijay-Kumar, M., Aitken, J.D., Carvalho, F.A., Cullender, T.C., Mwangi, S., Srinivasan, S., Sitaraman, S.V., Knight, R., Ley, R.E., and Gewirtz, A.T. (2010). Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. Science *328*, 228–231.

Wacklin, P., Mäkivuokko, H., Alakulppi, N., Nikkilä, J., Tenkanen, H., Räbinä, J., Partanen, J., Aranko, K., and Mättö, J. (2011). Secretor genotype (FUT2 gene) is strongly associated with the composition of Bifidobacteria in the human intestine. PLoS ONE *6*, e20113.

Walter, J., and Ley, R. (2011). The human gut microbiome: ecology and recent evolutionary changes. Annu. Rev. Microbiol. *65*, 411–429.

Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.Y., Keilbaugh, S.A., Bewtra, M., Knights, D., Walters, W.A., Knight, R., et al. (2011). Linking long-term dietary patterns with gut microbial enterotypes. Science *334*, 105–108.

Yang, J., Loos, R.J., Powell, J.E., Medland, S.E., Speliotes, E.K., Chasman, D.I., Rose, L.M., Thorleifsson, G., Steinthorsdottir, V., Mägi, R., et al. (2012). FTO genotype is associated with phenotypic variability of body mass index. Nature *490*, 267–272.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., et al. (2012). Human gut microbiome viewed across age and geography. Nature *486*, 222–227.

Zoetendal, E.G., Akkermans, A.D.L., Akkermans-van Vliet, W.M., Visser, J.A.G.M.d., and Vos, W.M.d. (2001). The host genotype affects the bacterial community in the human gastrointestinal tract. Microb. Ecol. Health Dis. *13*, 129–134.

BIBLIOGRAPHY

Aagaard, K., Ma, J., Antony, K. M., Ganu, R., Petrosino, J., and Versalovic, J. The placenta harbors a unique microbiome. *Science Translational Medicine*, 6(237), 2014.

Achenbach, P., Bonifacio, E., Koczwara, K., and Ziegler, A. G. Natural history of type 1 diabetes. *Diabetes*, 54:S25–S31, 2005.

Anand, P. K., Malireddi, R. K. S., Lukens, J. R., Vogel, P., Bertin, J., Lamkanfi, M., and Kanneganti, T.-D. Nlrp6 negatively regulates innate immunity and host defence against bacterial pathogens. *Nature*, 488(7411):389–+, 2012.

Andrew, T., Aviv, A., Falchi, M., Surdulescu, G. L., Gardner, J. P., Lu, X., Kimura, M., Kato, B. S., Valdes, A. M., and Spector, T. D. Mapping genetic loci that determine leukocyte telomere length in a large sample of unselected female sibling pairs. *American Journal of Human Genetics*, 78(3):480–486, 03 2006.

Antonioli, L., Pacher, P., Vizi, E. S., and Haskó, G. Cd39 and cd73 in immunity and inflammation. *Trends in molecular medicine*, 19(6): 355–367, 06 2013.

Arora, T., Loo, R. L., Anastasovska, J., Gibson, G. R., Tuohy, K. M., Sharma, R. K., Swann, J. R., Deaville, E. R., Sleeth, M. L., Thomas, E. L., Holmes, E., Bell, J. D., and Frost, G. Differential effects of two fermentable carbohydrates on central appetite regulation and body composition. *Plos One*, 7(8), 2012.

Arpaia, N., Campbell, C., Fan, X., Dikiy, S., van der Veeken, J., deRoos, P., Liu, H., Cross, J. R., Pfeffer, K., Coffer, P. J., and Rudensky, A. Y.

Metabolites produced by commensal bacteria promote peripheral regulatory t-cell generation. *Nature*, 504(7480):451–+, 2013.

Aubin, H.-J., Farley, A., Lycett, D., Lahmek, P., and Aveyard, P. Weight gain in smokers after quitting cigarettes: meta-analysis. *British Medical Journal*, 345, 2012.

Backhed, F., Ding, H., Wang, T., Hooper, L. V., Koh, G. Y., Nagy, A., Semenkovich, C. F., and Gordon, J. I. The gut microbiota as an environmental factor that regulates fat storage. *Proceedings of the National Academy of Sciences of the United States of America*, 101(44): 15718–15723, 2004.

Barness, L. A., Opitz, J. M., and Gilbert-Barness, E. Obesity: Genetic, molecular, and environmental aspects. *American Journal of Medical Genetics Part A*, 143A(24):3016–3034, 2007.

Barrios, C., Beaumont, M., Pallister, T., Villar, J., Goodrich, J. K., Clark, A., Pascual, J., Ley, R. E., Spector, T. D., Bell, J. T., and Menni, C. Gut-microbiota-metabolite axis in early renal function decline. *Plos One*, 10(8), 2015.

Bates, D., Maechler, M., Bolker, B., and Walker, S. Linear mixed-effects models using eigen and s4. 2014.

Beldi, G., Wu, Y., Banz, Y., Nowak, M., Miller, L., Enjyoji, K., Haschemi, A., Yegutkin, G. G., Candinas, D., Exley, M., and Robson, S. C. Natural killer t cell dysfunction in cd39-null mice protects against concanavalin a–induced hepatitis. *Hepatology (Baltimore, Md.)*, 48(3):841–852, 09 2008.

Benson, A. K., Kelly, S. A., Legge, R., Ma, F., Low, S. J., Kim, J., Zhang, M., Oh, P. L., Nehrenberg, D., Hua, K., Kachman, S. D., Moriyama, E. N., Walter, J., Peterson, D. A., and Pomp, D. Individuality in gut microbiota composition is a complex polygenic trait shaped by

multiple environmental and host genetic factors. *Proceedings of the National Academy of Sciences of the United States of America*, 107(44): 18933–18938, 2010.

Biedermann, L., Zeitz, J., Mwinyi, J., Sutter-Minder, E., Rehman, A., Ott, S. J., Steurer-Stey, C., Frei, A., Frei, P., Scharl, M., Loessner, M. J., Vavricka, S. R., Fried, M., Schreiber, S., Schuppler, M., and Rogler, G. Smoking cessation induces profound changes in the composition of the intestinal microbiota in humans. *Plos One*, 8(3), 2013.

Bonnet, M., Buc, E., Sauvanet, P., Darcha, C., Dubois, D., Pereira, B., Dechelotte, P., Bonnet, R., Pezet, D., and Darfeuille-Michaud, A. Colonization of the human gut by *E. coli* and colorectal cancer risk. *Clinical Cancer Research*, 20(4):859–867, 2014.

Borsellino, G., Kleinewietfeld, M., Di Mitri, D., Sternjak, A., Diamantini, A., Giometto, R., Höpner, S., Centonze, D., Bernardi, G., Dell'Acqua, M. L., Rossini, P. M., Battistini, L., Rötzschke, O., and Falk, K. Expression of ectonucleotidase cd39 by foxp3+ treg cells: hydrolysis of extracellular atp and immune suppression. *Blood*, 110 (4):1225–1232, 04 2007.

Bramham, C. R. and Messaoudi, E. Bdnf function in adult synaptic plasticity: The synaptic consolidation hypothesis. *Progress in Neurobiology*, 76(2):99–125, 6 2005.

Brion, M.-J. A., Shakhbazov, K., and Visscher, P. M. Calculating statistical power in mendelian randomization studies. *International Journal of Epidemiology*, 42(5):1497–1501, 2013.

Bruce-Keller, A. J., Salbaum, J. M., Luo, M., Blanchard, E., Taylor, C. M., Welsh, D. A., and Berthoud, H.-R. Obese-type gut microbiota induce neurobehavioral changes in the absence of obesity. *Biological Psychiatry*, 77(7):607–615, 2015.

Caporaso, J. G., Lauber, C. L., Costello, E. K., Berg-Lyons, D., Gonzalez, A., Stombaugh, J., Knights, D., Gajer, P., Ravel, J., Fierer, N., Gordon, J. I., and Knight, R. Moving pictures of the human microbiome. *Genome Biology*, 12(5), 2011.

Chambers, J. C., Elliott, P., Zabaneh, D., Zhang, W., Li, Y., Froguel, P., Balding, D., Scott, J., and Kooner, J. S. Common genetic variation near mc4r is associated with waist circumference and insulin resistance. *Nat Genet*, 40(6):716–718, 06 2008.

Chaput, J.-P., Pérusse, L., Després, J.-P., Tremblay, A., and Bouchard, C. Findings from the quebec family study on the etiology of obesity: Genetics and environmental highlights. *Current Obesity Reports*, 3 (1):54–66, 2014.

Choo, J. M., Leong, L. E., and Rogers, G. B. Sample storage conditions significantly influence faecal microbiome profiles. *Scientific Reports*, 5:16350 EP –, 11 2015.

Claesson, M. J., Jeffery, I. B., Conde, S., Power, S. E., O'Connor, E. M., Cusack, S., Harris, H. M. B., Coakley, M., Lakshminarayanan, B., O'Sullivan, O., Fitzgerald, G. F., Deane, J., O'Connor, M., Harnedy, N., O'Connor, K., O'Mahony, D., van Sinderen, D., Wallace, M., Brennan, L., Stanton, C., Marchesi, J. R., Fitzgerald, A. P., Shanahan, F., Hill, C., Ross, R. P., and O'Toole, P. W. Gut microbiota composition correlates with diet and health in the elderly. *Nature*, 488(7410), 2012.

Clarke, S. F., Murphy, E. F., O'Sullivan, O., Lucey, A. J., Humphreys, M., Hogan, A., Hayes, P., O'Reilly, M., Jeffery, I. B., Wood-Martin, R., Kerins, D. M., Quigley, E., Ross, R. P., O'Toole, P. W., Molloy, M. G., Falvey, E., Shanahan, F., and Cotter, P. D. Exercise and associated dietary extremes impact on gut microbial diversity. *Gut*, 2014.

Clemente, J. C., Pehrsson, E. C., Blaser, M. J., Sandhu, K., Gao, Z., Wang, B., Magris, M., Hidalgo, G., Contreras, M., Noya-Alarcón, Ó., Lander, O., McDonald, J., Cox, M., Walter, J., Oh, P. L., Ruiz, J. F., Rodriguez, S., Shen, N., Song, S. J., Metcalf, J., Knight, R., Dantas, G., and Dominguez-Bello, M. G. The microbiome of uncontacted amerindians. *Science Advances*, 1(3):e1500183, 04 2015.

Costello, E. K., Lauber, C. L., Hamady, M., Fierer, N., Gordon, J. I., and Knight, R. Bacterial community variation in human body habitats across space and time. *Science*, 326(5960):1694–1697, 2009.

Cotillard, A., Kennedy, S. P., Kong, L. C., Prifti, E., Pons, N., Le Chatelier, E., Almeida, M., Quinquis, B., Levenez, F., Galleron, N., Gougis, S., Rizkalla, S., Batto, J.-M., Renault, P., consortium, A. M., Dore, J., Zucker, J.-D., Clement, K., Ehrlich, S. D., and consortium members, A. M. Dietary intervention impact on gut microbial gene richness. *Nature*, 500(7464):585–588, 08 2013.

Cottone, M., Rosselli, M., Orlando, A., Oliva, L., Puleo, A., Cappello, M., Traina, M., Tonelli, F., and Pagliaro, L. Smoking-habits and recurrence in crohns-disease. *Gastroenterology*, 106(3):643–648, 1994.

Daneman, D. Type 1 diabetes. *Lancet*, 367(9513):847–858, 2006.

Daniel, H., Gholami, A. M., Berry, D., Desmarchelier, C., Hahne, H., Loh, G., Mondot, S., Lepage, P., Rothballer, M., Walker, A., Bohm, C., Wenning, M., Wagner, M., Blaut, M., Schmitt-Kopplin, P., Kuster, B., Haller, D., and Clavel, T. High-fat diet alters gut microbiota physiology in mice. *ISME J*, 8(2):295–308, 02 2014.

Davenport, E. R., Mizrahi-Man, O., Michelini, K., Barreiro, L. B., Ober, C., and Gilad, Y. Seasonal variation in human gut microbiome composition. *Plos One*, 9(3), 2014.

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., Ling, A. V., Devlin, A. S., Varma, Y., Fischbach, M. A., Biddinger, S. B., Dutton, R. J., and Turnbaugh, P. J. Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, 505(7484):559–+, 2014.

DiGiulio, D. B. Diversity of microbes in amniotic fluid. *Seminars in Fetal and Neonatal Medicine*, 17(1):2–11, 2011.

Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., and Knight, R. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences of the United States of America*, 107(26):11971–11975, 2010.

Dominguez-Bello, M. G., De Jesus-Laboy, K. M., Shen, N., Cox, L. M., Amir, A., Gonzalez, A., Bokulich, N. A., Song, S. J., Hoashi, M., Rivera-Vinas, J. I., Mendez, K., Knight, R., and Clemente, J. C. Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. *Nat Med*, 22(3):250–253, 03 2016.

Duncan, S. H., Lobley, G. E., Holtrop, G., Ince, J., Johnstone, A. M., Louis, P., and Flint, H. J. Human colonic microbiota associated with diet, obesity and weight loss. *International Journal of Obesity*, 32(11), 2008.

Duncan, S. H., Belenguer, A., Holtrop, G., Johnstone, A. M., Flint, H. J., and Lobley, G. E. Reduced dietary intake of carbohydrates by obese subjects results in decreased concentrations of butyrate and butyrate-producing bacteria in feces. *Applied and Environmental Microbiology*, 73(4):1073–1078, 2007.

Eggesbo, M., Botten, G., Stigum, H., Nafstad, P., and Magnus, P. Is delivery by cesarean section a risk factor for food allergy? *Journal of Allergy and Clinical Immunology*, 112(2):420–426, 2003.

Eknoyan, G. Adolphe quetelet (1796-1874) - the average man and indices of obesity. *Nephrology Dialysis Transplantation*, 23(1):47–51, 2008.

Eu-Ahsunthornwattana, J., Howey, R. A., and Cordell, H. J. Accounting for relatedness in family-based association studies: application to genetic analysis workshop 18 data. *BMC proceedings*, 8(Suppl 1 Genetic Analysis Workshop 18Vanessa Olmo):S79–S79, 2014.

Fairweather-Tait, S. J., Skinner, J., Guile, G. R., Cassidy, A., Spector, T. D., and MacGregor, A. J. Diet and bone mineral density study in postmenopausal women from the twinsuk registry shows a negative association with a traditional english dietary pattern and a positive association with wine. *Am J Clin Nutr*, 94(5):1371–1375, Nov 2011.

Feng, X.-T., Wang, T.-Z., Leng, J., Chen, Y., Liu, J.-B., Liu, Y., and Wang, W.-J. Palmitate contributes to insulin resistance through downregulation of the src-mediated phosphorylation of akt in c2c12 myotubes. *Bioscience Biotechnology and Biochemistry*, 76(7):1356–1361, 2012.

Ferrer, M., Ruiz, A., Lanza, F., Haange, S.-B., Oberbach, A., Till, H., Bargiela, R., Campoy, C., Segura, M. T., Richter, M., von Bergen, M., Seifert, J., and Suarez, A. Microbiota from the distal guts of lean and obese adolescents exhibit partial functional redundancy besides clear differences in community structure. *Environmental Microbiology*, 15(1):211–226, 2013.

Fierer, N., Hamady, M., Lauber, C. L., and Knight, R. The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 105(46):17994–17999, 2008.

Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P., and Forano, E. Microbial degradation of complex carbohydrates in the gut. *Gut microbes*, 3(4):289–306, 2012.

Flores, G. E., Caporaso, J. G., Henley, J. B., Rideout, J. R., Domogala, D., Chase, J., Leff, J. W., Vazquez-Baeza, Y., Gonzalez, A., Knight, R., Dunn, R. R., and Fierer, N. Temporal variability is a personalized feature of the human microbiome. *Genome Biology*, 15(12), 2014.

Fontana, L., Eagon, J. C., Trujillo, M. E., Scherer, P. E., and Klein, S. Visceral fat adipokine secretion is associated with systemic inflammation in obese humans. *Diabetes*, 56(4):1010–1013, 2007.

Fox, C. S., Massaro, J. M., Hoffmann, U., Pou, K. M., Maurovich-Horvat, P., Liu, C.-Y., Vasan, R. S., Murabito, J. M., Meigs, J. B., Cupples, L. A., D'Agostino, S., Ralph B., and O'Donnell, C. J. Abdominal visceral and subcutaneous adipose tissue compartments - association with metabolic risk factors in the framingham heart study. *Circulation*, 116(1):39–48, 2007.

Frank, D. N., Amand, A. L. S., Feldman, R. A., Boedeker, E. C., Harpaz, N., and Pace, N. R. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences of the United States of America*, 104(34):13780–13785, 2007.

Frank, D. N., Robertson, C. E., Hamm, C. M., Kpadeh, Z., Zhang, T., Chen, H., Zhu, W., Sartor, R. B., Boedeker, E. C., Harpaz, N., Pace, N. R., and Li, E. Disease phenotype and genotype are associ-

ated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. *Inflammatory Bowel Diseases*, 17(1):179–184, 2011.

Frayling, T. M., Timpson, N. J., Weedon, M. N., Zeggini, E., Freathy, R. M., Lindgren, C. M., Perry, J. R. B., Elliott, K. S., Lango, H., Rayner, N. W., Shields, B., Harries, L. W., Barrett, J. C., Ellard, S., Groves, C. J., Knight, B., Patch, A.-M., Ness, A. R., Ebrahim, S., Lawlor, D. A., Ring, S. M., Ben-Shlomo, Y., Jarvelin, M.-R., Sovio, U., Bennett, A. J., Melzer, D., Ferrucci, L., Loos, R. J. F., Barroso, I., Wareham, N. J., Karpe, F., Owen, K. R., Cardon, L. R., Walker, M., Hitman, G. A., Palmer, C. N. A., Doney, A. S. F., Morris, A. D., Smith, G. D., Hattersley, A. T., McCarthy, M. I., and Wellcome Trust Case, C. A common variant in the fto gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*, 316 (5826):889–894, 2007.

Friswell, M. K., Gika, H., Stratford, I. J., Theodoridis, G., Telfer, B., Wilson, I. D., and McBain, A. J. Site and strain-specific variation in gut microbiota profiles and metabolism in experimental mice. *Plos One*, 5(1), 2010.

Gao, D., Nong, S., Huang, X., Lu, Y., Zhao, H., Lin, Y., Man, Y., Wang, S., Yang, J., and Li, J. The effects of palmitate on hepatic insulin resistance are mediated by nadph oxidase 3-derived reactive oxygen species through jnk and p38(mapk) pathways. *Journal of Biological Chemistry*, 285(39):29965–29973, 2010.

General, S. http://www.cdc.gov/tobacco/data_statistics/sgr/50th-anniversary/index.htm. 2014.

Girardin, S. E., Boneca, I. G., Viala, J., Chamaillard, M., Labigne, A., Thomas, G., Philpott, D. J., and Sansonetti, P. J. Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (mdp) detection. *Journal of Biological Chemistry*, 278(11):8869–8872, 2003.

Goedert, J. J., Jones, G., Hua, X., Xu, X., Yu, G., Flores, R., Falk, R. T., Gail, M. H., Shi, J., Ravel, J., and Feigelson, H. S. Investigation of the association between the fecal microbiota and breast cancer in postmenopausal women: a population-based case-control pilot study. *Journal of the National Cancer Institute*, 107(8), 2015.

Goek, O.-N., Döring, A., Gieger, C., Heier, M., Koenig, W., Prehn, C., Römisch-Margl, W., Wang-Sattler, R., Illig, T., Suhre, K., Sekula, P., Zhai, G., Adamski, J., Köttgen, A., and Meisinger, C. Serum metabolite concentrations and decreased gfr in the general population. *American Journal of Kidney Diseases*, 60(2):197–206, 8 2012.

Goodrich, J. K., Waters, J. L., Poole, A. C., Sutter, J. L., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J. T., Spector, T. D., Clark, A. G., and Ley, R. E. Human genetics shape the gut microbiome. *Cell*, 159(4):789–799, 2014.

Gophna, U., Sommerfeld, K., Gophna, S., Doolittle, W. F., and van Zanten, S. J. O. V. Inerences between tissue-associated intestinal microfloras of patients with crohn's disease and ulcerative colitis. *Journal of Clinical Microbiology*, 44(11):4136–4141, 2006.

Gorbach, S. L. Estrogens, breast cancer and intestinal flora. *Reviews of Infectious Diseases*, 6:S85–S90, 1984.

Greenblum, S., Turnbaugh, P. J., and Borenstein, E. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proceedings of the National Academy of Sciences of the United States of America*, 109(2):594–599, 2012.

Gronlund, M. M., Lehtonen, O. P., Eerola, E., and Kero, P. Fecal microflora in healthy infants born by different methods of delivery:

Permanent changes in intestinal flora after cesarean delivery. *Journal of Pediatric Gastroenterology and Nutrition*, 28(1):19–25, 1999.

Guillet, C., Masgrau, A., and Boirie, Y. Is protein metabolism changed with obesity? *Current Opinion in Clinical Nutrition and Metabolic Care*, 14(1):89–92, 2011.

Guo, L., Milburn, M. V., Ryals, J. A., Lonergan, S. C., Mitchell, M. W., Wulff, J. E., Alexander, D. C., Evans, A. M., Bridgewater, B., Miller, L., Gonzalez-Garay, M. L., and Caskey, C. T. Plasma metabolomic profiles enhance precision medicine for volunteers of normal health. *Proceedings of the National Academy of Sciences*, 112(35):E4901–E4910, 2015.

Guyonnet, D., Chassany, O., Ducrotte, P., Picard, C., Mouret, M., Mercier, C.-H., and Matuchansky, C. Effect of a fermented milk containing bifidobacterium animalis dn-173 010 on the health-related quality of life and symptoms in irritable bowel syndrome in adults in primary care: a multicentre, randomized, double-blind, controlled trial. *Alimentary Pharmacology and Therapeutics*, 26(3):475–486, 2007.

Hildebrandt, M. A., Hoffmann, C., Sherrill-Mix, S. A., Keilbaugh, S. A., Hamady, M., Chen, Y.-Y., Knight, R., Ahima, R. S., Bushman, F., and Wu, G. D. High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology*, 137(5):1716–1724, 2009.

Hill, A. M., LaForgia, J., Coates, A. M., Buckley, J. D., and Howe, P. R. C. Estimating abdominal adipose tissue with dxa and anthropometry. *Obesity*, 15(2):504–510, 2007.

Hoppa, M. B., Collins, S., Ramracheya, R., Hodson, L., Amisten, S., Zhang, Q., Johnson, P., Ashcroft, F. M., and Rorsman, P. Chronic

palmitate exposure inhibits insulin secretion by dissociation of ca2+ channels from secretory granules. *Cell Metabolism*, 10(6):455–465, 2009.

Huh, S. Y., Rifas-Shiman, S. L., Zera, C. A., Edwards, J. W. R., Oken, E., Weiss, S. T., and Gillman, M. W. Delivery by caesarean section and risk of obesity in preschool age children: a prospective cohort study. *Archives of Disease in Childhood*, 97(7):610–616, 2012.

Huse, S. M., Ye, Y., Zhou, Y., and Fodor, A. A. A core human microbiome as viewed through 16s rrna sequence clusters. *Plos One*, 7(6), 2012.

Huszar, D., Lynch, C. A., Fairchild-Huntress, V., Dunmore, J. H., Fang, Q., Berkemeier, L. R., Gu, W., Kesterson, R. A., Boston, B. A., Cone, R. D., Smith, F. J., Campfield, L. A., Burn, P., and Lee, F. Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell*, 88(1):131–141, 1 1997.

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., Creasy, H. H., Earl, A. M., FitzGerald, M. G., Fulton, R. S., Giglio, M. G., Hallsworth-Pepin, K., Lobos, E. A., Madupu, R., Magrini, V., Martin, J. C., Mitreva, M., Muzny, D. M., Sodergren, E. J., Versalovic, J., Wollam, A. M., Worley, K. C., Wortman, J. R., Young, S. K., Zeng, Q., Aagaard, K. M., Abolude, O. O., Allen-Vercoe, E., Alm, E. J., Alvarado, L., Andersen, G. L., Anderson, S., Appelbaum, E., Arachchi, H. M., Armitage, G., Arze, C. A., Ayvaz, T., Baker, C. C., Begg, L., Belachew, T., Bhonagiri, V., Bihan, M., Blaser, M. J., Bloom, T., Bonazzi, V., Brooks, J. P., Buck, G. A., Buhay, C. J., Busam, D. A., Campbell, J. L., Canon, S. R., Cantarel, B. L., Chain, P. S. G., Chen, I. M. A., Chen, L., Chhibba, S., Chu, K., Ciulla, D. M., Clemente, J. C., Clifton, S. W., Conlan, S., Crabtree, J., Cutting, M. A., Davidovics, N. J., Davis, C. C., DeSantis,

T. Z., Deal, C., Delehaunty, K. D., Dewhirst, F. E., Deych, E., Ding, Y., Dooling, D. J., Dugan, S. P., Dunne, W. M., Durkin, A. S., Edgar, R. C., Erlich, R. L., Farmer, C. N., Farrell, R. M., Faust, K., Feldgarden, M., Felix, V. M., Fisher, S., Fodor, A. A., Forney, L. J., Foster, L., Di Francesco, V., Friedman, J., Friedrich, D. C., Fronick, C. C., Fulton, L. L., Gao, H., Garcia, N., Giannoukos, G., Giblin, C., Giovanni, M. Y., Goldberg, J. M., Goll, J., Gonzalez, A., Griggs, A., et al. Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402):207–214, 2012.

Jamshidi, Y., Snieder, H., Wang, X., Spector, T. D., Carter, N. D., and O'Dell, S. D. Common polymorphisms in the socs3 gene are not associated with body weight, insulin sensitivity or lipid profile in normal female twins. *Diabetologia*, 49(2):306–310, 02 2006.

Jiménez, E., Fernández, L., Marín, M., Martín, R., Odriozola, J., Nueno-Palop, C., Narbad, A., Olivares, M., Xaus, J., and Rodríguez, J. Isolation of commensal bacteria from umbilical cord blood of healthy neonates born by cesarean section. *Current Microbiology*, 51 (4):270–274, 2005.

Karlsen, T. H., Franke, A., Melum, E., Kaser, A., Hov, J. R., Balschun, T., Lie, B. A., Bergquist, A., Schramm, C., Weismueller, T. J., Gotthardt, D., Rust, C., Philipp, E. E. R., Fritz, T., Henckaerts, L., Weersma, R. K., Stokkers, P., Ponsioen, C. Y., Wijmenga, C., Sterneck, M., Nothnagel, M., Hampe, J., Teufel, A., Runz, H., Rosenstiel, P., Stiehl, A., Vermeire, S., Beuers, U., Manns, M. P., Schrumpf, E., Boberg, K. M., and Schreiber, S. Genome-wide association analysis in primary sclerosing cholangitis. *Gastroenterology*, 138(3):1102–1111, 2010.

Karra, E., O'Daly, O. G., Choudhury, A. I., Yousseif, A., Millership, S., Neary, M. T., Scott, W. R., Chandarana, K., Manning, S., Hess, M. E.,

Iwakura, H., Akamizu, T., Millet, Q., Gelegen, C., Drew, M. E., Rahman, S., Emmanuel, J. J., Williams, S. C. R., Ruether, U. U., Bruening, J. C., Withers, D. J., Zelaya, F. O., and Batterham, R. L. A link between fto, ghrelin, and impaired brain food-cue responsivity. *Journal of Clinical Investigation*, 123(8):3539–3551, 2013.

Kelly, T., Yang, W., Chen, C. S., Reynolds, K., and He, J. Global burden of obesity in 2005 and projections to 2030. *International Journal of Obesity*, 32(9):1431–1437, 2008.

Khachatryan, Z. A., Ktsoyan, Z. A., Manukyan, G. P., Kelly, D., Ghazaryan, K. A., and Aminov, R. I. Predominant role of host genetics in controlling the composition of gut microbiota. *PLoS ONE*, 3(8): e3064, 2008.

Kilkkinen, A., Rissanen, H., Klaukka, T., Pukkala, E., Heliovaara, M., Huovinen, P., Mannisto, S., Aromaa, A., and Knekt, P. Antibiotic use predicts an increased risk of cancer. *International Journal of Cancer*, 123(9):2152–2155, 2008.

Klaus, B. and Strimmer., K. *fdrtool: Estimation of (Local) False Discovery Rates and Higher Criticism*, 2015.

Klöting, N., Graham, T. E., Berndt, J., Kralisch, S., Kovacs, P., Wason, C. J., Fasshauer, M., Schön, M. R., Stumvoll, M., Blüher, M., and Kahn, B. B. Serum retinol-binding protein is more highly expressed in visceral than in subcutaneous adipose tissue and is a marker of intra-abdominal fat mass. *Cell Metabolism*, 6(1):79–87, 7 2007.

Kostic, A. D., Gevers, D., Siljander, H., Vatanen, T., Hyotylainen, T., Hamalainen, A.-M., Peet, A., Tillmann, V., Poho, P., Mattila, I., Lahdesmaki, H., Franzosa, E. A., Vaarala, O., de Goffau, M., Harmsen, H., Ilonen, J., Virtanen, S. M., Clish, C. B., Oresic, M., Huttenhower, C., Knip, M., Xavier, R. J., and Grp, D. S. The dynamics

of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host & Microbe*, 17(2):260–273, 2015.

L, B., CB, F., P, Z., TM, R., A, L., and PA, D. Association of antibiotics in infancy with early childhood obesity. *JAMA Pediatrics*, 168(11): 1063–1069, 11 2014.

Lapchak, P. A. and Hefti, F. Bdnf and ngf treatment in lesioned rats - effects on cholinergic function and weight-gain. *Neuroreport*, 3(5): 405–408, 1992.

Larsen, L. H., Echwald, S. M., Sørensen, T. I. A., Andersen, T., Wulff, B. S., and Pedersen, O. Prevalence of mutations and functional analyses of melanocortin 4 receptor variants identified among 750 men with juvenile-onset obesity. *The Journal of Clinical Endocrinology & Metabolism*, 90(1):219–224, 2005.

Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N., and Smith, G. D. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine*, 27(8):1133–1163, 2008.

Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J.-M., Kennedy, S., Leonard, P., Li, J., Burgdorf, K., Grarup, N., Jorgensen, T., Brandslund, I., Nielsen, H. B., Juncker, A. S., Bertalan, M., Levenez, F., Pons, N., Rasmussen, S., Sunagawa, S., Tap, J., Tims, S., Zoetendal, E. G., Brunak, S., Clement, K., Dore, J., Kleerebezem, M., Kristiansen, K., Renault, P., Sicheritz-Ponten, T., de Vos, W. M., Zucker, J.-D., Raes, J., Hansen, T., Bork, P., Wang, J., Ehrlich, S. D., Pedersen, O., and Meta, H. I. T. C. Richness of human gut microbiome correlates with metabolic markers. *Nature*, 500(7464):541–+, 2013.

Levey, A. S., Stevens, L. A., Schmid, C. H., Zhang, Y., Castro, I., Alejandro F., Feldman, H. I., Kusek, J. W., Eggers, P., Van Lente, F., Greene, T., Coresh, J., and Chronic Kidney Dis Epidemiology, C. A new equation to estimate glomerular filtration rate. *Annals of Internal Medicine*, 150(9):604–612, 2009.

Ley, R. E., Backhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., and Gordon, J. I. Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences of the United States of America*, 102 (31), 2005.

Ley, R. E., Peterson, D. A., and Gordon, J. I. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*, 124(4):837–848, 2006.

Lindberg, E., Tysk, C., Andersson, K., and Jarnerot, G. Smoking and inflammatory bowel disease - a case control study. *Gut*, 29(3):352–357, 1988.

Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., Powell, C., Vedantam, S., Buchkovich, M. L., Yang, J., Croteau-Chonka, D. C., Esko, T., Fall, T., Ferreira, T., Gustafsson, S., Kutalik, Z., Luan, J. a., Magi, R., Randall, J. C., Winkler, T. W., Wood, A. R., Workalemahu, T., Faul, J. D., Smith, J. A., Hua Zhao, J., Zhao, W., Chen, J., Fehrmann, R., Hedman, A. K., Karjalainen, J., Schmidt, E. M., Absher, D., Amin, N., Anderson, D., Beekman, M., Bolton, J. L., Bragg-Gresham, J. L., Buyske, S., Demirkan, A., Deng, G., Ehret, G. B., Feenstra, B., Feitosa, M. F., Fischer, K., Goel, A., Gong, J., Jackson, A. U., Kanoni, S., Kleber, M. E., Kristiansson, K., Lim, U., Lotay, V., Mangino, M., Mateo Leach, I., Medina-Gomez, C., Medland, S. E., Nalls, M. A., Palmer, C. D., Pasko, D., Pechlivanis, S., Peters, M. J., Prokopenko, I., Shungin, D., Stancakova, A., Strawbridge, R. J., Ju Sung, Y., Tanaka, T., Teumer, A., Trompet, S., van der

Laan, S. W., van Setten, J., Van Vliet-Ostaptchouk, J. V., Wang, Z., Yengo, L., Zhang, W., Isaacs, A., Albrecht, E., Arnlov, J., Arscott, G. M., Attwood, A. P., Bandinelli, S., Barrett, A., Bas, I. N., Bellis, C., Bennett, A. J., Berne, C., Blagieva, R., Bluher, M., Bohringer, S., Bonnycastle, L. L., Bottcher, Y., Boyd, H. A., Bruinenberg, M., Caspersen, I. H., Ida Chen, Y.-D., Clarke, R., Warwick Daw, E., de Craen, A. J. M., Delgado, G., Dimitriou, M., et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*, 518(7538):197–206, 2015.

Loos, R. J. F., Lindgren, C. M., Li, S., Wheeler, E., Zhao, J. H., Prokopenko, I., Inouye, M., Freathy, R. M., Attwood, A. P., Beckmann, J. S., Berndt, S. I., The Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, Bergmann, S., Bennett, A. J., Bingham, S. A., Bochud, M., Brown, M., Cauchi, S., Connell, J. M., Cooper, C., Smith, G. D., Day, I., Dina, C., De, S., Dermitzakis, E. T., Doney, A. S. F., Elliott, K. S., Elliott, P., Evans, D. M., Farooqi, I. S., Froguel, P., Ghori, J., Groves, C. J., Gwilliam, R., Hadley, D., Hall, A. S., Hattersley, A. T., Hebebrand, J., Heid, I. M., KORA, Herrera, B., Hinney, A., Hunt, S. E., Jarvelin, M.-R., Johnson, T., Jolley, J. D. M., Karpe, F., Keniry, A., Khaw, K.-T., Luben, R. N., Mangino, M., Marchini, J., McArdle, W. L., McGinnis, R., Meyre, D., Munroe, P. B., Morris, A. D., Ness, A. R., Neville, M. J., Nica, A. C., Ong, K. K., O'Rahilly, S., Owen, K. R., Palmer, C. N. A., Papadakis, K., Potter, S., Pouta, A., Qi, L., Study, N. H., Randall, J. C., Rayner, N. W., Ring, S. M., Sandhu, M. S., Scherag, A., Sims, M. A., Song, K., Soranzo, N., Speliotes, E. K., Initiative, D. G., Syddall, H. E., Teichmann, S. A., Timpson, N. J., Tobias, J. H., Uda, M., Study, T. S., Vogel, C. I. G., Wallace, C., Waterworth, D. M., Weedon, M. N., Consortium, T. W. T. C. C., Willer, C. J., FUSION, Wraight, V. L., Yuan, X., Zeggini, E., Hirschhorn, J. N., Strachan, D. P., Ouwehand,

W. H., Caulfield, M. J., Samani, N. J., Frayling, T. M., Vollenweider, P., Waeber, G., Mooser, V., Deloukas, P., McCarthy, M. I., Wareham, N. J., Barroso, I., Jacobs, K. B., Chanock, S. J., Hayes, R. B., Lamina, C., Gieger, C., Illig, T., Meitinger, T., Wichmann, H.-E., Kraft, P., Hankinson, S. E., Hunter, D. J., Hu, F. B., Lyon, H. N., Voight, B. F., Ridderstrale, M., Groop, L., Scheet, P., Sanna, S., Abecasis, G. R., Albai, G., Nagaraja, R., Schlessinger, D., Jackson, A. U., Tuomilehto, J., Collins, F. S., Boehnke, M., and Mohlke, K. L. Common variants near mc4r are associated with fat mass, weight and risk of obesity. *Nature genetics*, 40(6):768–775, 06 2008.

Macfabe, D. F. Short-chain fatty acid fermentation products of the gut microbiome: implications in autism spectrum disorders. *Microbial ecology in health and disease*, 23, 2012.

Major, G. and Spiller, R. Irritable bowel syndrome, inflammatory bowel disease and the microbiome. *Current Opinion in Endocrinology Diabetes and Obesity*, 21(1):15–21, 2014.

Mangino, M., Richards, J. B., Soranzo, N., Zhai, G., Aviv, A., Valdes, A. M., Samani, N. J., Deloukas, P., and Spector, T. D. A genome-wide association study identifies a novel locus on chromosome 18q12.2 influencing white cell telomere length. *Journal of Medical Genetics*, 46(7):451–454, 07 2009.

Marsh, H., O'Shea, T. J., and Reynolds, J. E. *Ecology and Conservation of the Sirenia*. Conservation Biology, 2012.

Mason, M. R., Preshaw, P. M., Nagaraja, H. N., Dabdoub, S. M., Rahman, A., and Kumar, P. S. The subgingival microbiome of clinically healthy current and never smokers. *Isme Journal*, 9(1):268–272, 2015.

Meijer, K., de Vos, P., and Priebe, M. G. Butyrate and other short-chain fatty acids as modulators of immunity: what relevance for

health? *Current Opinion in Clinical Nutrition and Metabolic Care*, 13 (6):715–721, 2010.

Mekkes, M. C., Weenen, T. C., Brummer, R. J., and Claassen, E. The development of probiotic treatment in obesity: a review. *Beneficial Microbes*, 5(1):19–28, 2014.

Menni, C., Fauman, E., Erte, I., Perry, J. R. B., Kastenmueller, G., Shin, S.-Y., Petersen, A.-K., Hyde, C., Psatha, M., Ward, K. J., Yuan, W., Milburn, M., Palmer, C. N. A., Frayling, T. M., Trimmer, J., Bell, J. T., Gieger, C., Mohney, R. P., Brosnan, M. J., Suhre, K., Soranzo, N., and Spector, T. D. Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach. *Diabetes*, 62 (12):4270–4276, 2013a.

Menni, C., Kastenmüller, G., Petersen, A. K., Bell, J. T., Psatha, M., Tsai, P.-C., Gieger, C., Schulz, H., Erte, I., John, S., Brosnan, M. J., Wilson, S. G., Tsaprouni, L., Lim, E. M., Stuckey, B., Deloukas, P., Mohney, R., Suhre, K., Spector, T. D., and Valdes, A. M. Metabolomic markers reveal novel pathways of ageing and early development in human populations. *International Journal of Epidemiology*, 07 2013b.

Moran, J. P., Walter, J., Tannock, G. W., Tonkonogy, S. L., and Sartor, R. B. Bifidobacterium animalis causes extensive duodenitis and mild colonic inflammation in monoassociated interleukin-lodeficient mice. *Inflammatory Bowel Diseases*, 15(7):1022–1031, 2009.

Morotomi, M., Nagai, F., and Watanabe, Y. Description of christensenella minuta gen. nov., sp nov., isolated from human faeces, which forms a distinct branch in the order clostridiales, and proposal of christensenellaceae fam. nov. *International Journal of Systematic and Evolutionary Microbiology*, 62:144–149, 2012.

Morris, A., Beck, J. M., Schloss, P. D., Campbell, T. B., Crothers, K., Curtis, J. L., Flores, S. C., Fontenot, A. P., Ghedin, E., Huang, L., Jablonski, K., Kleerup, E., Lynch, S. V., Sodergren, E., Twigg, H., Young, V. B., Bassis, C. M., Venkataraman, A., Schmidt, T. M., Weinstock, G. M., and Lung, H. I. V. Microbiome Project. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *American Journal of Respiratory and Critical Care Medicine*, 187 (10):1067–1075, 2013.

Nalls, M. A., Couper, D. J., Tanaka, T., van Rooij, F. J. A., Chen, M.-H., Smith, A. V., Toniolo, D., Zakai, N. A., Yang, Q., Greinacher, A., Wood, A. R., Garcia, M., Gasparini, P., Liu, Y., Lumley, T., Folsom, A. R., Reiner, A. P., Gieger, C., Lagou, V., Felix, J. F., Völzke, H., Gouskova, N. A., Biffi, A., Döring, A., Völker, U., Chong, S., Wiggins, K. L., Rendon, A., Dehghan, A., Moore, M., Taylor, K., Wilson, J. G., Lettre, G., Hofman, A., Bis, J. C., Pirastu, N., Fox, C. S., Meisinger, C., Sambrook, J., Arepalli, S., Nauck, M., Prokisch, H., Stephens, J., Glazer, N. L., Cupples, L. A., Okada, Y., Takahashi, A., Kamatani, Y., Matsuda, K., Tsunoda, T., Tanaka, T., Kubo, M., Nakamura, Y., Yamamoto, K., Kamatani, N., Stumvoll, M., Tönjes, A., Prokopenko, I., Illig, T., Patel, K. V., Garner, S. F., Kuhnel, B., Mangino, M., Oostra, B. A., Thein, S. L., Coresh, J., Wichmann, H.-E., Menzel, S., Lin, J., Pistis, G., Uitterlinden, A. G., Spector, T. D., Teumer, A., Eiriksdottir, G., Gudnason, V., Bandinelli, S., Frayling, T. M., Chakravarti, A., van Duijn, C. M., Melzer, D., Ouwehand, W. H., Levy, D., Boerwinkle, E., Singleton, A. B., Hernandez, D. G., Longo, D. L., Soranzo, N., Witteman, J. C. M., Psaty, B. M., Ferrucci, L., Harris, T. B., O'Donnell, C. J., and Ganesh, S. K. Multiple loci are associated with white blood cell phenotypes. *PLoS Genet*, 7(6): e1002113, 06 2011.

Natividad, J. M. M., Petit, V., Huang, X., de Palma, G., Jury, J., Sanz, Y., Philpott, D., Rodenas, C. L. G., McCoy, K. D., and Verdu, E. F. Commensal and probiotic bacteria influence intestinal barrier function and susceptibility to colitis in nod1-/-;nod2-/- mice. *Inflammatory Bowel Diseases*, 18(8):1434–1446, 2012.

Neale, M. C. and Cardon, L. R. *Methodology for genetic studies of twins and families.*, volume 67. Springer Netherlands, 1992.

Negele, K., Heinrich, J., Borte, M., von Berg, A., Schaaf, B., Lehmann, I., Wichmann, H. E., Bolte, G., and Grp, L. S. Mode of delivery and development of atopic disease during the first 2 years of life. *Pediatric Allergy and Immunology*, 15(1):48–54, 2004.

Niewczas, M. A., Sirich, T. L., Mathew, A. V., Skupien, J., Mohney, R. P., Warram, J. H., Smiles, A., Huang, X., Walker, W., Byun, J., Karoly, E. D., Kensicki, E. M., Berry, G. T., Bonventre, J. V., Pennathur, S., Meyer, T. W., and Krolewski, A. S. Uremic solutes and risk of end-stage renal disease in type 2 diabetes: metabolomic study. *Kidney Int*, 85(5):1214–1224, 05 2014.

Nishimura, J., Masaki, T., Arakawa, M., Seike, M., and Yoshimatsu, H. Isoleucine prevents the accumulation of tissue triglycerides and upregulates the expression of pparα and uncoupling protein in diet-induced obese mice. *The Journal of Nutrition*, 140(3):496–500, 2010.

Ogura, Y., Bonen, D. K., Inohara, N., Nicolae, D. L., Chen, F. F., Ramos, R., Britton, H., Moran, T., Karaliuskas, R., Duerr, R. H., Achkar, J. P., Brant, S. R., Bayless, T. M., Kirschner, B. S., Hanauer, S. B., Nunez, G., and Cho, J. H. A frameshift mutation in nod2 associated with susceptibility to crohn's disease. *Nature*, 411(6837):603–606, 2001.

Oikari, L. E., Okolicsanyi, R. K., Qin, A., Yu, C., Griffiths, L. R., and Haupt, L. M. Cell surface heparan sulfate proteoglycans as novel

markers of human neural stem cell fate determination. *Stem Cell Research*, 16(1):92–104, 1 2016.

O'Keefe, S. J. D., Ou, J., Aufreiter, S., O'Connor, D., Sharma, S., Sepulveda, J., Fukuwatari, T., Shibata, K., and Mawhinney, T. Products of the colonic microbiota mediate the effects of diet on colon cancer risk. *Journal of Nutrition*, 139(11):2044–2048, 2009.

Patti, M. E. and Kahn, B. B. Nutrient sensor links obesity with diabetes risk. *Nature Medicine*, 10(10):1049–1050, 2004.

Payne, A. N., Chassard, C., Zimmermann, M., Mueller, P., Stinca, S., and Lacroix, C. The metabolic activity of gut microbiota in obese children is increased compared with normal-weight children and exhibits more exhaustive substrate utilization. *Nutrition & Diabetes*, 1, 2011.

Petnicki-Ocwieja, T., Hrncir, T., Liu, Y.-J., Biswas, A., Hudcovic, T., Tlaskalova-Hogenova, H., and Kobayashi, K. S. Nod2 is required for the regulation of commensal microbiota in the intestine. *Proceedings of the National Academy of Sciences of the United States of America*, 106(37), 2009.

Philippe, D., Favre, L., Foata, F., Adolfsson, O., Perruisseau-Carrier, G., Vidal, K., Reuteler, G., Dayer-Schneider, J., Mueller, C., and Blum, S. *Bifidobacterium lactis* attenuates onset of inflammation in a murine model of colitis. *World Journal of Gastroenterology*, 17(4): 459–469, 2011.

Pihlajamaki, J., Gylling, H., Miettinen, T. A., and Laakso, M. Insulin resistance is associated with increased cholesterol synthesis and decreased cholesterol absorption in normoglycemic men. *Journal of Lipid Research*, 45(3):507–512, 2004.

Pimentel, M., Gunsalus, R. P., Rao, S. S., and Zhang, H. Methanogens in human health and disease. *Am J Gastroenterol Suppl*, 1(1):28–33, 07 2012.

Preidis, G. A., Keaton, M. A., Campeau, P. M., Bessard, B. C., Conner, M. E., and Hotez, P. J. The undernourished neonatal mouse metabolome reveals evidence of liver and biliary dysfunction, inflammation, and oxidative stress. *Journal of Nutrition*, 144(3):273–281, 2014.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H. B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Dore, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Bork, P., Ehrlich, S. D., Wang, J., and MetaHIT Consortium. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285):59–U70, 2010.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.

Ravel, J., Gajer, P., Abdo, Z., Schneider, G. M., Koenig, S. S. K., McCulle, S. L., Karlebach, S., Gorle, R., Russell, J., Tacket, C. O., Brotman, R. M., Davis, C. C., Ault, K., Peralta, L., and Forney, L. J. Vaginal microbiome of reproductive-age women. *Proceedings of the National Academy of Sciences of the United States of America*, 108:4680–4687, 2011.

Rehman, A., Sina, C., Gavrilova, O., Haesler, R., Ott, S., Baines, J. F., Schreiber, S., and Rosenstiel, P. Nod2 is essential for temporal devel-

opment of intestinal microbial communities. *Gut*, 60(10):1354–1362, 2011.

Rice, T., Daw, E. W., Gagnon, J., Bouchard, C., Leon, A. S., Skinner, J. S., Wilmore, J. H., and Rao, D. C. Familial resemblance for body composition measures: The heritage family study. *Obesity Research*, 5(6):557–562, 1997.

Rios, M. Bdnf and the central control of feeding: accidental bystander or essential player? *Trends in Neurosciences*, 36(2):83–90, 2013.

Roederer, M., Quaye, L., Mangino, M., Beddall, M. H., Mahnke, Y., Chattopadhyay, P., Tosi, I., Napolitano, L., Barberio, M. T., Menni, C., Villanova, F., Meglio, P. D., Spector, T. D., and Nestle, F. O. The genetic architecture of the human immune system: A bioresource for autoimmunity and disease pathogenesis. *Cell*, 161(2):387 – 403, 2015.

Roediger, W. E. W. and Babidge, W. Human colonocyte detoxification. *Gut*, 41(6):731–734, 1997.

Romero-Corral, A., Somers, V. K., Sierra-Johnson, J., Thomas, R. J., Collazo-Clavell, M. L., Korinek, J., Allison, T. G., Batsis, J. A., Sert-Kuniyoshi, F. H., and Lopez-Jimenez, F. Accuracy of body mass index in diagnosing obesity in the adult general population. *International Journal of Obesity*, 32(6):959–966, 2008.

Rooks, M. G., Veiga, P., Wardwell-Scott, L. H., Tickle, T., Segata, N., Michaud, M., Gallini, C. A., Beal, C., van Hylckama-Vlieg, J. E. T., Ballal, S. A., Morgan, X. C., Glickman, J. N., Gevers, D., Huttenhower, C., and Garrett, W. S. Gut microbiome composition and function in experimental colitis during active disease and treatment-induced remission. *Isme Journal*, 8(7):1403–1417, 2014.

Saez-Lara, M. J., Gomez-Llorente, C., Plaza-Diaz, J., and Gil, A. The role of probiotic lactic acid bacteria and bifidobacteria in the prevention and treatment of inflammatory bowel disease and other related diseases: a systematic review of randomized human clinical trials. *BioMed research international*, 2015:505878–505878, 2015.

Salminen, S., Gibson, G. R., McCartney, A. L., and Isolauri, E. Influence of mode of delivery on gut microbiota composition in seven year old children. *Gut*, 53(9):1388–1389, 2004.

Samuel, B. S. and Gordon, J. I. A humanized gnotobiotic mouse model of host-archaeal-bacterial mutualism. *Proceedings of the National Academy of Sciences of the United States of America*, 103(26): 10011–10016, 2006.

Saric, J., Wang, Y., Li, J., Coen, M., Utzinger, J., Marchesi, J. R., Keiser, J., Veselkov, K., Lindon, J. C., Nicholson, J. K., and Holmes, E. Species variation in the fecal metabolome gives insight into differential gastrointestinal function. *Journal of Proteome Research*, 7(1): 352–360, 2008.

Scherag, A., Jarick, I., Grothe, J., Biebermann, H., Scherag, S., Volckmar, A.-L., Vogel, C. I. G., Greene, B., Hebebrand, J., and Hinney, A. Investigation of a genome wide association signal for obesity: Synthetic association and haplotype analyses at the melanocortin 4 receptor gene locus. *PLoS ONE*, 5(11):e13967, 11 2010.

Schnorr, S. L., Candela, M., Rampelli, S., Centanni, M., Consolandi, C., Basaglia, G., Turroni, S., Biagi, E., Peano, C., Severgnini, M., Fiori, J., Gotti, R., De Bellis, G., Luiselli, D., Brigidi, P., Mabulla, A., Marlowe, F., Henry, A. G., and Crittenden, A. N. Gut microbiome of the hadza hunter-gatherers. *Nature Communications*, 5, 2014.

Schwiertz, A., Taras, D., Schaefer, K., Beijer, S., Bos, N. A., Donus, C., and Hardt, P. D. Microbiota and scfa in lean and overweight healthy subjects. *Obesity*, 18(1):190–195, 2010.

Sergentanis, T. N., Zagouri, F., and Zografos, G. C. Is antibiotic use a risk factor for breast cancer? a meta-analysis. *Pharmacoepidemiology and Drug Safety*, 19(11):1101–1107, 2010.

Shen, Q., Zhao, L., and Tuohy, K. M. High-level dietary fibre up-regulates colonic fermentation and relative abundance of saccharolytic bacteria within the human faecal microbiota in vitro. *European Journal of Nutrition*, 51(6):693–705, 2012.

Shin, S.-Y., Fauman, E. B., Petersen, A.-K., Krumsiek, J., Santos, R., Huang, J., Arnold, M., Erte, I., Forgetta, V., Yang, T.-P., Walter, K., Menni, C., Chen, L., Vasquez, L., Valdes, A. M., Hyde, C. L., Wang, V., Ziemek, D., Roberts, P., Xi, L., Grundberg, E., Waldenberger, M., Richards, J. B., Mohney, R. P., Milburn, M. V., John, S. L., Trimmer, J., Theis, F. J., Overington, J. P., Suhre, K., Brosnan, M. J., Gieger, C., Kastenmueller, G., Spector, T. D., Soranzo, N., and Multiple Tissue Human Expression, R. An atlas of genetic influences on human blood metabolites. *Nature Genetics*, 46(6):543–550, 2014.

Sironi, A. M., Petz, R., De Marchi, D., Buzzigoli, E., Ciociaro, D., Positano, V., Lombardi, M., Ferrannini, E., and Gastaldelli, A. Impact of increased visceral and cardiac fat on cardiometabolic risk and disease. *Diabetic Medicine*, 29(5):622–627, 2012.

Smemo, S., Tena, J. J., Kim, K.-H., Gamazon, E. R., Sakabe, N. J., Gomez-Marin, C., Aneas, I., Credidio, F. L., Sobreira, D. R., Wasserman, N. F., Lee, J. H., Puviindran, V., Tam, D., Shen, M., Son, J. E., Vakili, N. A., Sung, H.-K., Naranjo, S., Acemel, R. D., Manzanares, M., Nagy, A., Cox, N. J., Hui, C.-C., Gomez-Skarmeta, J. L., and

Nobrega, M. A. Obesity-associated variants within fto form long-range functional connections with irx3. *Nature*, 507(7492):371–375, 03 2014.

Snijder, M. B., Visser, M., Dekker, J. M., Seidell, J. C., Fuerst, T., Tylavsky, F., Cauley, J., Lang, T., Nevitt, M., and Harris, T. B. The prediction of visceral fat by dual-energy x-ray absorptiometry in the elderly: a comparison with computed tomography and anthropometry. *International Journal of Obesity*, 26(7):984–993, 2002.

Snijder, M. B., Visser, M., Dekker, J. M., Goodpaster, B. H., Harris, T. B., Kritchevsky, S. B., De Rekeneire, N., Kanaya, A. M., Newman, A. B., Tylavsky, F. A., and Seidell, J. C. Low subcutaneous thigh fat is a risk factor for unfavourable glucose and lipid levels, independently of high abdominal fat. the health abc study. *Diabetologia*, 48(2):301–308, 2005.

Song, S. J., Lauber, C., Costello, E. K., Lozupone, C. A., Humphrey, G., Berg-Lyons, D., Caporaso, J. G., Knights, D., Clemente, J. C., Nakielny, S., Gordon, J. I., Fierer, N., and Knight, R. Cohabiting family members share microbiota with one another and with their dogs. *Elife*, 2, 2013.

Speakman, J. R., Rance, K. A., and Johnstone, A. M. Polymorphisms of the fto gene are associated with variation in energy intake, but not energy expenditure. *Obesity*, 16(8):1961–1965, 2008.

Speliotes, E. K., Willer, C. J., Berndt, S. I., Monda, K. L., Thorleifsson, G., Jackson, A. U., Allen, H. L., Lindgren, C. M., Luan, J., Maegi, R., Randall, J. C., Vedantam, S., Winkler, T. W., Qi, L., Workalemahu, T., Heid, I. M., Steinthorsdottir, V., Stringham, H. M., Weedon, M. N., Wheeler, E., Wood, A. R., Ferreira, T., Weyant, R. J., Segre, A. V., Estrada, K., Liang, L., Nemesh, J., Park, J.-H., Gustafsson, S., Kilpelaenen, T. O., Yang, J., Bouatia-Naji, N., Esko, T., Feitosa, M. F.,

Kutalik, Z., Mangino, M., Raychaudhuri, S., Scherag, A., Smith, A. V., Welch, R., Zhao, J. H., Aben, K. K., Absher, D. M., Amin, N., Dixon, A. L., Fisher, E., Glazer, N. L., Goddard, M. E., Heard-Costa, N. L., Hoesel, V., Hottenga, J.-J., Johansson, A., Johnson, T., Ketkar, S., Lamina, C., Li, S., Moffatt, M. F., Myers, R. H., Narisu, N., Perry, J. R. B., Peters, M. J., Preuss, M., Ripatti, S., Rivadeneira, F., Sandholt, C., Scott, L. J., Timpson, N. J., Tyrer, J. P., van Wingerden, S., Watanabe, R. M., White, C. C., Wiklund, F., Barlassina, C., Chasman, D. I., Cooper, M. N., Jansson, J.-O., Lawrence, R. W., Pellikka, N., Prokopenko, I., Shi, J., Thiering, E., Alavere, H., Alibrandi, M. T. S., Almgren, P., Arnold, A. M., Aspelund, T., Atwood, L. D., Balkau, B., Balmforth, A. J., Bennett, A. J., Ben-Shlomo, Y., Bergman, R. N., Bergmann, S., Biebermann, H., Blakemore, A. I. F., Boes, T., Bonnycastle, L. L., Bornstein, S. R., Brown, M. J., Buchanan, T. A., et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics*, 42(11):937–U53, 2010.

Stappenbeck, T. S., Hooper, L. V., and Gordon, J. I. Developmental regulation of intestinal angiogenesis by indigenous microbes via paneth cells. *Proceedings of the National Academy of Sciences of the United States of America*, 99(24):15451–15455, 2002.

Stewart, J. A., Chadwick, V. S., and Murray, A. Investigations into the influence of host genetics on the predominant eubacteria in the faecal microflora of children. *Journal of Medical Microbiology*, 54(12): 1239–1242, 2005.

Suez, J., Korem, T., Zeevi, D., Zilberman-Schapira, G., Thaiss, C. A., Maza, O., Israeli, D., Zmora, N., Gilad, S., Weinberger, A., Kuperman, Y., Harmelin, A., Kolodkin-Gal, I., Shapiro, H., Halpern, Z., Segal, E., and Elinav, E. Artificial sweeteners induce glucose in-

tolerance by altering the gut microbiota. *Nature*, 514(7521):181–+, 2014.

Sun, X., Feng, R., Li, Y., Lin, S., Zhang, W., Li, Y., Sun, C., and Li, S. Histidine supplementation alleviates inflammation in the adipose tissue of high-fat diet-induced obese rats via the nf-κb- and pparγ-involved pathways. *British Journal of Nutrition*, 112:477–485, 8 2014.

Tamim, H. M., Hanley, J. A., Hajeer, A. H., Boivin, J.-F., and Collet, J.-P. Risk of breast cancer in relation to antibiotic use. *Pharmacoepidemiology and Drug Safety*, 17(2):144–150, 2008.

Tamim, H. M., Hajeer, A. H., Boivin, J.-F., and Collet, J.-P. Association between antibiotic use and risk of prostate cancer. *International Journal of Cancer*, 127(4):952–960, 2010.

Teucher, B., Skinner, J., Skidmore, P. M. L., Cassidy, A., Fairweather-Tait, S. J., Hooper, L., Roe, M. A., Foxall, R., Oyston, S. L., Cherkas, L. F., Perks, U. C., Spector, T. D., and MacGregor, A. J. Dietary patterns and heritability of food choice in a uk female twin cohort. *Twin Research and Human Genetics*, 10(5):734–748, 2007.

Thavagnanam, S., Fleming, J., Bromley, A., Shields, M. D., and Cardwell, C. R. A meta-analysis of the association between caesarean section and childhood asthma. *Clinical & Experimental Allergy*, 38 (4):629–633, 2008.

Thibault, R., Blachier, F., Darcy-Vrillon, B., de Coppet, P., Bourreille, A., and Segain, J.-P. Butyrate utilization by the colonic mucosa in inflammatory bowel diseases: A transport deficiency. *Inflammatory Bowel Diseases*, 16(4):684–695, 2010.

Thomson, C. A. Diet and breast cancer: Understanding risks and benefits. *Nutrition in Clinical Practice*, 27(5):636–650, 2012.

Thrush, A. B., Heigenhauser, G. J., Mullen, K. L., Wright, D. C., and Dyck, D. J. Palmitate acutely induces insulin resistance in isolated muscle from obese but not lean humans. *American Journal of Physiology-Regulatory Integrative and Comparative Physiology*, 294 (4):R1205–R1212, 2008.

Trasande, L., Blustein, J., Liu, M., Corwin, E., Cox, L. M., and Blaser, M. J. Infant antibiotic exposures and early-life body mass. *Int J Obes*, 37(1):16–23, 01 2013.

Trompette, A., Gollwitzer, E. S., Yadava, K., Sichelstiel, A. K., Sprenger, N., Ngom-Bru, C., Blanchard, C., Junt, T., Nicod, L. P., Harris, N. L., and Marsland, B. J. Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nature Medicine*, 20(2):159–166, 2014.

Tsukinowa, E., Karita, S., Asano, S., Wakai, Y., Oka, Y., Furuta, M., and Goto, M. Fecal microbiota of a dugong (dugong dugong) in captivity at toba aquarium. *Journal of General and Applied Microbiology*, 54 (1):25–38, 2008.

Turnbaugh, P. J. and Gordon, J. I. The core gut microbiome, energy balance and obesity. *Journal of Physiology-London*, 587(17):4153–4158, 2009.

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., and Gordon, J. I. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, 444(7122):1027–1031, 2006.

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., and Gordon, J. I. The human microbiome project. *Nature*, 449 (7164):804–810, 2007.

Turnbaugh, P. J., Baeckhed, F., Fulton, L., and Gordon, J. I. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host & Microbe*, 3(4):213–223, 2008.

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., Sogin, M. L., Jones, W. J., Roe, B. A., Affourtit, J. P., Egholm, M., Henrissat, B., Heath, A. C., Knight, R., and Gordon, J. I. A core gut microbiome in obese and lean twins. *Nature*, 457 (7228):480–U7, 2009a.

Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R., and Gordon, J. I. The effect of diet on the human gut microbiome: A metagenomic analysis in humanized gnotobiotic mice. *Science Translational Medicine*, 1(6), 2009b.

Tzatsos, A. and Kandror, K. V. Nutrients suppress phosphatidylinositol 3-kinase/akt signaling via raptor-dependent mtor-mediated insulin receptor substrate 1 phosphorylation. *Molecular and Cellular Biology*, 26(1):63–76, 2006.

Vaahtovuo, J., Toivanen, P., and Eerola, E. Bacterial composition of murine fecal microflora is indigenous and genetically guided. *FEMS Microbiology Ecology*, 44(1):131–136, 05 2003.

van Buuren, S. and Groothuis-Oudshoorn, K. mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45 (3):1–67, 2011.

Vink, J., Willemsen, G., and Boomsma, D. Heritability of smoking initiation and nicotine dependence. *Behavior Genetics*, 35(4):397–406, 2005.

Walters, W. A., Xu, Z., and Knight, R. Meta-analyses of human gut microbes associated with obesity and ibd. *Febs Letters*, 588(22):4223–4233, 2014.

Wang, J., Tang, H., Zhang, C., Zhao, Y., Derrien, M., Rocher, E., Vlieg, J. E. T. V.-H., Strissel, K., Zhao, L., Obin, M., and Shen, J. Modulation of gut microbiota during probiotic-mediated attenuation of metabolic syndrome in high fat diet-fed mice. *Isme Journal*, 9(1): 1–15, 2015.

Wang, J.-L., Chang, C.-H., Lin, J.-W., Wu, L.-C., Chuang, L.-M., and Lai, M.-S. Infection, antibiotic therapy and risk of colorectal cancer: A nationwide nested case-control study in patients with type 2 diabetes mellitus. *International Journal of Cancer*, 135(4):956–967, 2014.

Wang, T. J., Larson, M. G., Vasan, R. S., Cheng, S., Rhee, E. P., McCabe, E., Lewis, G. D., Fox, C. S., Jacques, P. F., Fernandez, C., O'Donnell, C. J., Carr, S. A., Mootha, V. K., Florez, J. C., Souza, A., Melander, O., Clish, C. B., and Gerszten, R. E. Metabolite profiles and the risk of developing diabetes. *Nat Med*, 17(4):448–453, 04 2011.

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., and Parkinson, H. The nhgri gwas catalog, a curated resource of snp-trait associations. *Nucleic Acids Research*, 42(D1):D1001–D1006, 2014.

Wen, L., Ley, R. E., Volchkov, P. Y., Stranges, P. B., Avanesyan, L., Stonebraker, A. C., Hu, C., Wong, F. S., Szot, G. L., Bluestone, J. A., Gordon, J. I., and Chervonsky, A. V. Innate immunity and intestinal microbiota in the development of type 1 diabetes. *Nature*, 455(7216): 1109–U10, 2008.

Wijeyesekera, A., Clarke, P. A., Bictash, M., Brown, I. J., Fidock, M., Ryckmans, T., Yap, I. K. S., Chan, Q., Stamler, J., Elliott, P., Holmes, E., and Nicholson, J. K. Quantitative uplc-ms/ms analysis of the gut microbial co-metabolites phenylacetylglutamine, 4-cresyl sulphate and hippurate in human urine: Intermap study. *Analytical methods : advancing methods and applications*, 4(1):65–72, 01 2012.

Willett, W. C. Dietary fats and coronary heart disease. *Journal of Internal Medicine*, 272(1):13–24, 2012.

Wolever, T. M. S., Spadafora, P. J., Cunnane, S. C., and Pencharz, P. B. Propionate inhibits incorporation of colonic 1,2-c-13 acetate into plasma-lipids in humans. *American Journal of Clinical Nutrition*, 61 (6):1241–1247, 1995.

Xia, J. and Wishart, D. S. Msea: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Research*, 38:W71–W77, 2010.

Yadav, H., Jain, S., and Sinha, P. R. Antidiabetic effect of probiotic dahi containing lactobacillus acidophilus and lactobacillus casei in high fructose fed rats. *Nutrition*, 23(1):62–68, 2007.

Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R. N., Anokhin, A. P., Heath, A. C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J. G., Lozupone, C. A., Lauber, C., Clemente, J. C., Knights, D., Knight, R., and Gordon, J. I. Human gut microbiome viewed across age and geography. *Nature*, 486(7402):222–+, 2012.

Yusuf, S., Hawken, S., Ôunpuu, S., Dans, T., Avezum, A., Lanas, F., McQueen, M., Budaj, A., Pais, P., Varigos, J., and Lisheng, L. Effect of potentially modifiable risk factors associated with myocardial

infarction in 52 countries (the interheart study): case-control study. *The Lancet*, 364(9438):937–952, 2004.

Zhai, G., Teumer, A., Stolk, L., Perry, J. R. B., Vandenput, L., Coviello, A. D., Koster, A., Bell, J. T., Bhasin, S., Eriksson, J., Eriksson, A., Ernst, F., Ferrucci, L., Frayling, T. M., Glass, D., Grundberg, E., Haring, R., Hedman, Å. K., Hofman, A., Kiel, D. P., Kroemer, H. K., Liu, Y., Lunetta, K. L., Maggio, M., Lorentzon, M., Mangino, M., Melzer, D., Miljkovic, I., Nica, A., Penninx, B. W. J. H., Vasan, R. S., Rivadeneira, F., Small, K. S., Soranzo, N., Uitterlinden, A., Völzke, H., Wilson, S. G., Xi, L., Zhuang, W. V., Harris, T. B., Murabito, J. M., Ohlsson, C., Murray, A., de Jong, F. H., Spector, T. D., Wallaschofski, H., and Consortium, M. Eight common genetic variants associated with serum dheas levels suggest a key role in ageing mechanisms. *PLoS Genet*, 7(4):e1002025 EP –, 04 2011.

Zhang, C., Zhang, M., Wang, S., Han, R., Cao, Y., Hua, W., Mao, Y., Zhang, X., Pang, X., Wei, C., Zhao, G., Chen, Y., and Zhao, L. Interactions between gut microbiota, host genetics and diet relevant to development of metabolic syndromes in mice. *Isme Journal*, 4(2), 2010.

Zhang, H., DiBaise, J. K., Zuccolo, A., Kudrna, D., Braidotti, M., Yu, Y., Parameswaran, P., Crowell, M. D., Wing, R., Rittmann, B. E., and Krajmalnik-Brown, R. Human gut microbiota in obesity and after gastric bypass. *Proceedings of the National Academy of Sciences of the United States of America*, 106(7), 2009.

Zhou, X. and Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, 44(7):821–U136, 2012.

Zimmer, J., Lange, B., Frick, J. S., Sauer, H., Zimmermann, K., Schwiertz, A., Rusch, K., Klosterhalfen, S., and Enck, P. A vegan

or vegetarian diet substantially alters the human colonic faecal microbiota. *European Journal of Clinical Nutrition*, 66(1):53–60, 2012.

Zoetendal, E. G., Akkermans, A. D. L., Akkermans-van Vliet, W. M., de Visser, J. A. G. M., and de Vos, W. M. The host genotype affects the bacterial community in the human gastrointestinal tract. *Microbial Ecology in Health and Disease*, 13(3):129–134, 2001.